Université de Montréal

Influences génétiques et environnementales sur la variabilité et l'unicité des activations cérébrales chez l'humain:
Un devis familial de jumeaux sur la base de données d'imagerie cérébrale du Human Connectome Project

par

Yassine Benhajali

Département d'Anthropologie, Faculté d'Arts et Sciences
Faculté des arts et des sciences

Thèse présentée en vue de l'obtention du grade de
Philosophiæ Doctor (Ph.D.)
en Anthropologie

September 16, 2021

Université de Montréal

Faculté des arts et des sciences

Cette thèse intitulée

Influences génétiques et environnementales sur la variabilité et l'unicité des activations cérébrales chez l'humain:
Un devis familial de jumeaux sur la base de données d'imagerie cérébrale du Human Connectome Project

présentée par

Yassine Benhajali

a été évaluée par un jury composé des personnes suivantes :

Bernard Bernier
(président-rapporteur)

Daniel Pérusse
(directeur de recherche)

Pierre Bellec
(codirecteur)

Guy Lanoue
(membre du jury)

Erin W. Dickie
(examinatrice externe)

(représentant du doyen de la FESP)

Résumé

Le comportement humain est à la fois singulier et universel. La singularité serait principalement due aux trajectoires de vie propre à chaque individu (variant entre autres selon leur culture) alors que l'universalité émanerait d'une nature universelle ancrée dans un génome universel. Démêler les influences de la nature et de la culture sur le comportement humain est le Saint Graal de l'anthropologie biologique. J'aborde cette question en explorant les effets génétiques et environnementaux sur les bases psychiques du comportement. Plus particulièrement, je teste l'hypothèse que la singularité et l'universalité comportementales humaines s'observent au plan psychique par l'exploration de leur substrat neurobiologique, et que ce substrat possède à la fois un ancrage génétique et environnemental. À l'aide de données d'imagerie par résonance magnétique fonctionnelle (IRMf) recueillies auprès de 862 participants du Human Connectome Project (HCP), j'analyse les activations cérébrales liées à sept tâches socio-cognitives qui recoupent diverses facultés, dont le langage, la mémoire, la prise de risque, la logique, les émotions, la motricité et le raisonnement social. Après avoir groupé les sujets selon la similarité de leurs patrons d'activation cérébrale (c.-à-d. leurs sous-types neurobiologiques), j'estime l'influence génétique et environnementale sur la variabilité interindividuelle de ces divers sous-types. Les résultats démontrent bel et bien l'existence d'un regroupement des sujets selon la similarité de leurs cartes d'activation cérébrale lors d'une même tâche socio-cognitive, ce qui reflète à la fois le caractère singulier et universel des corrélats neuronaux d'un comportement observable. La variabilité interindividuelle constatée dans ces regroupements cérébraux témoigne quant à elle d'effets génétiques (héritabilité) ainsi qu'environnementaux (environnementalité), dont les ampleurs respectives varient selon la nature de la tâche effectuée. De plus, les sous-types cérébraux mis à jour révèlent une association avec les mesures comportementales et de performance effectuées lors des diverses tâches à l'étude. Enfin, les sous-types neurobiologiques résultant des diverses tâches partagent certaines bases génétiques. Dans leur ensemble, ces résultats appuient la notion que le comportement humain, ainsi que les processus neurobiologiques le sous-tendant, sont des phénotypes au même titre qu'un caractère morphologique ou physiologique, c'est-à-dire qu'ils sont le résultat de l'expression conjointe de bases génétiques (nature) et environnementales (culture).

Mots-clés : Comportement, universalité, variabilité, corrélat neurobiologique, imagerie cérébrale fonctionnelle, génétique quantitative.

Abstract

Human behaviour is both singular and universal. Singularity is believed to be mainly due to life trajectories unique to each individual (influenced among others by culture), whereas universality would stem from a universal nature resulting from a panhuman genome. Unravelling the influences of nature and nurture on human behaviour is the Holy Grail of biological anthropology. I approach this issue by exploring genetic and environmental influences on the neuropsychological underpinnings of behaviour. In particular, I test the hypothesis that the singularity and universality of human behaviour are also observed at the psychological level through the exploration of the neurobiological basis of behaviour, and that these bases have both genetic and environmental sources. Using Functional Magnetic Resonance Imaging (fMRI) data of 862 participants from the Human Connectome Project (HCP), I analyze brain activation related to 7 socio-cognitive tasks covering language, memory, risk taking, logic, emotions, motor skills, and social reasoning. After grouping subjects according to the similarity of their brain activation patterns (neurobiological subtypes), I estimate the genetic and environmental influences on the variation between participants on these subtypes. The inter-individual variability in cerebral groupings appears to have both genetic (heritability) and environmental (environmentality) sources that vary according to the particular psychological task involved. Moreover, these neurobiological subtypes show an association with behavioural and performance measures assessed by the socio-cognitive tasks. Finally, the neurobiological subtypes across the 7 tasks share common genetic links. Overall, the results support the notion that human behaviour, as well as its underlying neurobiological processes, are phenotypes in the same way as morphology or physiology, i.e., are the results of the integrated expression of a genetic basis (nature) and environmental influences (nurture).

Keywords : Behaviour, universality, variability, neurobiological correlate, functional brain imaging, quantitative genetics.

Table des matières

Résumé 3

Abstract 5

Liste des tableaux 11

Liste des figures 12

Liste des sigles et des abréviations 19

Remerciements 20

Chapitre 1. Introduction 21

1.1. Le comportement humain : définitions 22

1.2. Neurosciences du comportement humain 23

1.2.1. IRMf 23

1.3. Universalité et variabilité en IRMf du comportement 25

1.4. Objectifs spécifiques 28

Chapitre 2. Méthodes et concepts dans l'analyse des phénotypes neurocomportementaux 31

2.1. La collecte de données 31

2.2. L'analyse des données 34

2.3. Les sources de variabilité liées à l'acquisition et le traitement des données en IRMf 36

2.4. Génétique quantitative du comportement humain 38

First Article. A standardized protocol for efficient and reliable quality control of brain registration in functional MRI studies 42

1. Introduction 43

1.1. Brain registration 43
1.2. Visual QC 44
1.3. Inter-rater agreement 44
1.4. Crowdsourced QC 45

2. Method 47
2.1. QC protocol building 47
2.2. Brain slices 49
2.3. Anatomical landmarks 49
2.4. Rating guidelines 49
2.5. Zooniverse platform 50
2.6. Brain images validation and training sets 50
2.7. Raters 51
2.8. Agreement statistics 52
2.9. Consensus panels 53

3. Results 54
3.1. Expert raters achieved moderate agreement, with "Fail" rating being the most reliable 54
3.2. Large panels of expert or zooniverse raters give convergent, reliable QC ratings 55
3.3. Small consensus panels of expert (N=3) or Zooniverse (N=20) raters achieve reliable QC ratings 56

4. Discussion 56
4.1. Visual QC 57
4.2. Inter-rater Agreement 57
4.3. Crowd sourced QC 57
4.4. Limitations of the study 58
4.5. Future work: impact of QC on downstream analyses 59

5. Conclusion 60

6. Contribution to the Field Statement 60

7. Acknowledgement 61

Supplementary material 62

Second Article. Subtypes of brain activation are heritable and genetically linked with behaviour in the Human Connectome Project sample 64

1. Introduction 65
1.1. Inter-individual differences in brain activation 66
1.2. Brain subtypes 66
1.3. Genetic of brain activation subtypes 67
1.4. Aims and hypotheses 68

2. Method 69
2.1. Subjects 69
2.2. Imaging data 69
2.3. Image preprocessing 70
2.4. Registration and motion quality control 71
2.5. Task fMRI 71
2.5.1. Working Memory 71
2.5.2. Incentive Processing (gambling) 72
2.5.3. Motor 72
2.5.4. Language Processing 72
2.5.5. Social Cognition (Theory of Mind) 72
2.5.6. Relational Processing 72
2.5.7. Emotion Processing 73
2.6. Activation maps 73
2.7. Subtypes of activation maps and weights extraction 73
2.8. Clustering Behavioural measures 74
2.9. Phenotype definitions for heritability and pleiotropy estimates 76
2.10. Univariate analysis of additive genetic variance 76
2.11. Bivariate genetic analyses 77

3. Results 77
3.1. Activation maps for each of the 7 tasks have good face validity are consistent with previous studies 77
3.2. Individual activation maps cluster in subgroups with similar spatial distribution 78
3.2.1. Language and working-memory subtypes weights associate with Behavioural phenotypes 80

3.3. Heritability of subtypes weights ranges from weak to strong 81
3.4. In-scanner and Out-of-scanner behavioural measures are heritable 81
3.5. Subtypes weights are genetically associated with behaviour 84

4. Discussion 85
4.1. Brain subtypes are associated with Behavioural measures 85
4.2. Heritability of brain subtypes and Behavioural measure 87
4.3. Genetic correlation between brain subtypes and Behavioural phenotypes 88
4.4. Cognitive factors underlying brain subtypes? 89
4.5. Limitations 89
4.6. Future directions 90

Supplementary material 92

Third Article. Two highly pleiotropic mirror subtypes of brain activation shared across tasks in the Human Connectome Project 106

1. Introduction 107
1.1. Inter-individual differences in brain activation and brain activation subtypes 108
1.2. Genetic correlation 109
1.3. Aims and hypotheses 109

2. Method 110
2.1. Subjects 110
2.2. Imaging data 110
2.3. Image preprocessing 111
2.4. Registration and motion quality control 111
2.5. Activation maps 112
2.6. Meta-subtypes from the 7 HCP tasks 112
2.7. Meta-subtype generation and association with behavioural measures 114

3. Results 115
3.1. Subtypes are genetically correlated across tasks and cluster into meta-subtypes 117
3.2. Genetic meta-subtypes are associated with Behavioural measures 118

4. Discussion 119
4.1. Genetic correlation between subtypes 119
4.2. Across tasks meta-subtypes 120
4.3. Association with Behavioural measures 120

4.4. Limitations . . . 121
4.5. Future directions . . . 121
5. Conclusions . . . 122
Supplementary material . . . 123
Task fMRI . . . 123
Working Memory . . . 123
Incentive Processing (gambling) . . . 123
Motor . . . 123
Language Processing . . . 123
Social Cognition (Theory of Mind) . . . 124
Relational Processing . . . 124
Emotion Processing . . . 124
Spatial meta-subtypes . . . 127
Spatial meta-subtypes generation . . . 127
Meta-subtypes spatially overlap with resting state networks . . . 127
Chapitre 3. Discussion . . . 130
Réduction de la variabilité liée au prétraitement des données en IRMf . . . 131
Les sous-types d'activation cérébrale ont un ancrage génétique et associé avec le comportement . . . 132
Deux grands modes de variation génétiques influencent les mesures comportementales dans notre échantillon. . . . 133
Interdisciplinarité neuroscience, anthropologie et génétique . . . 134
Références bibliographiques . . . 135

Liste des tableaux

1 **Objectifs -** Le tableau ci-dessous résume les objectifs spécifiques de chaque chapitre : 30

1 **Coefficients de la variance dans la covariance des apparentés** (Tiré du livre de Falconer et Mackay 1996) 40

2 **Reported agreement in visual inspection of sMRI data on QC studies -** QC studies since 2010 that uses sMRI and reported their inter/intra-raters agreement (Dadar et al., 2018; Esteban et al., 2017; Fonov et al., 2018; Klapwijk et al., 2019; Pizarro et al., 2016; Rosen et al., 2018) 46

3 **In-scanner task performance description -** In-scanner accuracy and reaction time measures used in this study. 93

Liste des figures

1 **Le domaine spatio-temporel en neurosciences et les principales méthodes disponibles pour l'étude du système nerveux en 2014** (tiré de Sejnowski et collègues 2014). Chaque région colorée représente le domaine spatio-temporel de chaque méthode d'imagerie fonctionnelle du cerveau jusqu'en 2014. Les régions vides illustrent l'utilisation de techniques de mesure, et les régions pleines de méthodes de perturbation (principalement invasives). En bas à droite, une représentation des méthodes disponibles en 1988 fait remarquer le vide que l'IRMf est venue combler en moins de deux décennies. **EEG** électroencéphalographie ; **MEG**, magnétoencéphalographie ; **TEP**, tomographie par émission de positrons ; **VSD**, colorant sensible au voltage ; **TMS**, stimulation magnétique transcrânienne ; **2-DG**, 2-désoxy-glucose. 24

2 **Variabilité en connectivité fonctionnelle et expansion corticale évolutive. A** - Expansion corticale entre un macaque et un humain. **B** - Variabilité fonctionnelle inter-sujets. **C** - Corrélation élevée (r = 0,52) entre l'expansion corticale et la variabilité fonctionnelle. (Mueller et al., 2013) 26

1 **Nombre de publications utilisant l'International Neuroimaging Data-sharing Initiative - INDI** (tiré de Milham et collègues, 2018). Le nombre de publications scientifiques recensées entre 2010 et 2016 dans des revues à révision par les pairs dans lesquelles les différentes bases de données partagées de INDI ont été utilisées. **CORR**: Consortium for Reliability and Reproducibility, **ABIDE**: Autism Brain Imaging Data Exchange, **NKI**: Nathan Kline Institute-Rockland Sample, **ADHD**: Attention Deficit Hyperactivity Disorder Sample, et **FCP**: 1000 Functional Connectomes Project. 32

2 **Le potentiel de réutilisation de différents types de partage de données**. Plus le coût et le nombre de données à partager sont élevés, plus le potentiel d'utilisation par la communauté est élevé. Tirée du texte de Poldrack et Gorgolewski 2014 33

3 **Carte d'activation cérébrale d'analyse de groupe**. Les zones cérébrales activées durant cette tâche représentent les régions connues pour le traitement de l'information auditive et du langage. Tiré de Barch et collègues 2013. 35

4 **Effet des artefacts sur les images BOLD** - **G1**: Distorsion spatiale - **G2**: Perte de signal - **G3**: Images fantômes. Figure tirée de Soares 2016. 36

5 **Contrôle de qualité visuel de l'alignement des images T1 dans l'espace de référence**. Les flèches jaunes indiquent les endroits où l'alignement a échoué. **a**: Sujet avec un alignement qui a échoué de manière claire. **b**: Sujet avec un alignement qui a échoué de manière subtile. 37

6 **QC protocol for brain registration- a: brain slices** - he rater is presented with two sets of brain slices (3 axial, 3 sagittal and 3 coronal), one of them showing the template in stereotactic space and the other showing an individual T1 brain after registration. In the interface, the two images are superimposed and the rater can flip between them to visually assess the registration. **b: anatomical landmarks** - The landmarks for QC included: the outline of the brain (A), tentorium cerebelli (B), cingulate sulcu s(C), parieto-cingulate sulcus occipital fissure (D), calcarine fissure (E), the lateral ventricles (F), central sulcus (G) and the hippocampal formation (H) bilaterally. The landmarks were outlined in stereotaxic space. **c: rating guidelines** - The boundaries of red landmarks act as "confidence interval" for registration: an area is tagged as a misregistration only if the target structure falls outside the boundaries. **d: tags** - Raters put tags on each misregistered brain structure. **e: final rating** - A final decision is reached on the quality of registration: an image with no tags is rated OK, one or more nonadjacent tags are rated Maybe, two or more adjacent tags are rated Fail. An image that is excessively blurry is also rated Fail. 48

7 **Between-expert agreement - a**: Matrix of Kappa agreement between raters (top). Note that R1 to R9 are identification codes for the different expert raters. The distribution of agreement is also presented (bottom). For example, the boxplot for R1 shows the agreement between R1 and R2-R9. **b-d**: Matrix and distribution for the Dice agreement between raters in the OK (**b**), Maybe (**c**) and Fail (**d**) categories. 54

8 **Zooniverse, expert and radiologist agreements - a**: Matrix of Kappa agreement between consensus of experts (Ec), zooniverse users (Zc) and radiologist (Ra) raters, in rows, vs individual experts (R1-R9), in column (top). The

distribution of agreement is also presented (bottom). **b-d**: Matrix and distribution for the Dice agreement in the OK (b), Maybe (c) and Fail (d) categories. 55

9 **Agreement between small panels of raters for both experts and Zooniverse panels - a**: Matrix of Kappa agreement between large panel consensus of experts (Ec), zooniverse users (Zc) and a small panel of expert (Ec1 = 3 rater, Ec2 = 3 rater, Ec3 = 3 rater) and small panel of Zooniverse raters (Zc1 = 20 rater, Zc2 = 21 rater) (top). The distribution of agreement is also presented (bottom). **b-d**: Dice distribution between group consensus in the OK (**b**), Maybe (**c**) and Fail (**d**) categories. 56

10 **Distribution of images as a function of the number of ratings (mean of 21.76 ± 2.75, range 15 to 27).** 62

11 **Brain Match tutorial interface - a**: A tutorial popup when first accessing the brain match interface showing the main goal of the task. **b**: A panel of annotated examples of what is a good and bad match between highlighted brain areas. More examples follows on how to tag images and how to decide the final rating of each image (for more detailed tutorial visit brain-match) 63

12 **Activation subtypes maps generation and weights extraction- a**. From each subject, an un-thresholded statistical parametric map (SPM) is vectorized then stacked. **b**. Nuisance regression from stacked SPM (FD: Frame Displacement, Sys: Systolic blood pressure, Dias: Diastolic blood pressure, BMI: Body Mass Index). **c**. From the residual stack, a subject by subject correlation matrix is clustered into subgroups of activation maps (subtypes). **d**. Each subtype is represented spatially by a mean activation map of all subjects within the same cluster . **e**. Each subtype is represented on the subject level by its weight (correlation between a vectorized subject activation map and a vectorized subtype map). **f**. Individual weights are correlated genetically (ρ_g) and phenotypically (ρ_p) with behavioural measures. 74

13 **Behavioural dimensionality reduction- a**. A correlation matrix of a collection of 74 behavioural tests are clustered using k-means. The Silhouette plot shows 4 local maxima of cluster separation. The 13 clusters cut off are **Speed**: task reaction time measures, **Physio**: physiological measures, **Mix2** and **Mix1**: a mixed category of measures, **Executive**: working memory and cognitive flexibility measures, **AcadSkil**: language and vocabulary test along with academic completion measure, **FluInt**: fluid intelligence measures, **SelfReg**: working

memory and cognitive flexibility measures, **PosAff**: positive affect measures, **NegAff**: negative affect measures, **DisorInt**: internalized disorders, **Somatic**: symptoms, problems or sensory manifestations related to the body, **DisorExt**: dysregulated externalized behaviour. 75

14 **Subtypes of activation maps - a: Language subtypes** - The grand mean for the language statistical parametric maps revealed a cluster of high activation in the superior temporal gyrus during the "Story minus Math" condition. Subtype maps 1 to 5 (S1-5) was defined as the average map of each subgroup or cluster. Subtypes show systematic deviation from the grand mean that range from a predominantly 'deactivating' pattern toward a more 'activating' pattern of brain activity **b: Working memory subtypes** - The grand mean for the working memory statistical parametric maps revealed a cluster of high activation in the shows bilateral activation in the superior frontal gyrus, middle frontal gyrus, inferior frontal gyrus and the inferior parietal lobule during the "0-Back minus 2-Back" condition. Subtype maps 1 to 5 (S1-5) was defined as the average map of each subgroup or cluster. They show activation and deactivation patterns along the superior frontal gyrus (BA 6), middle frontal gyrus (BA 9 and 10), inferior frontal gyrus (BA 47) and the inferior parietal lobule (BA 40). 79

15 **Phenotypic correlation between brain subtypes and behavioural measures - a: Correlation between language subtypes and in-scanner task performance.** The highest correlation estimates are between the language task accuracy (LanAcc) measure and subtypes weights (S1-S5). See supplementary material Table 3 for a list of abbreviations used for the in-scanner task performance. **b - Correlation between language subtypes and out-of-scanner behavioural measures:** The highest correlation estimates are between the academic skills (AcadSkil) behavioural cluster and subtypes weights. See supplementary material figure Sup 19 for a list of abbreviations used for the behavioural clusters. **c - Correlation between working memory subtypes and in-scanner task performance.** The highest correlation estimates are between the working memory task accuracy measures and all subtypes weights. **d - Correlation between working memory subtypes and out-of-scanner behavioural measures.** The highest correlation estimates are between the academic skills (AcadSkil) behavioural cluster and all subtypes weights. 80

16 **Heritability estimate - a: Similarity between activation maps by relationship.** Average correlation of brain activation maps between pairs of MZ twins, DZ twins, siblings and unrelated individuals. The magnitude of correlations varied substantially across task domains, but a dosage effect was observed in almost every task, with correlation for MZ twins greater than DZ twins, greater than siblings or unrelated individuals. **b: Heritability of subtype weights for all HCP tasks.** Significant heritability was identified for the 7 HCP tasks: gambling task (W: Win, L: Loss, W-L: Win vs Loss), language task (S-M: Story vs Math), motor task (LH: Left Hand, LF: Left Foot, T: Tong, RF: Right Foot, LF: Left Foot), social task (Mt: Mental, Mt-R: Mental vs Random), working memory task (2B: 2-Back, 0B: 0-Back, 2B-0B: 2-Back vs 0-Back). Relational task (M: Match, R: Relational, M-R: Match vs Relational) and the emotional task (F-N: Fear vs Neutral). Each task condition was represented by 5 subtypes' weights. Heritability estimate (H2r) on these subtype weights ranged from 0 to 0.62. **c: Heritability of in-scanner task performance.** Heritability ranged from 0.12 to 0.57 with the highest estimates for working memory and relational accuracy scores. See supplementary material Table 1 for a detailed description of in-scanner behavioural labels. **d: Heritability of out-of-scanner behavioural clusters.** Heritability ranged from 0.15 to 0.75 with the highest estimate for academic skills (AcadSkil). See supplementary material figure Sup 19 for a list of abbreviations used for the clustered behavioural measures. 82

17 **Genetic correlation between brain phenotypes and behavioural phenotypes. a - Genetic correlation between language subtypes and in-scanner behavioural tests:** The highest genetic associations are between the working memory reaction time and the five language subtypes weights. See supplementary material Table 1 for a list of abbreviations used for the in-scanner task performance. **b - Genetic correlation between language subtypes and out-of-scanner behavioural measures.** All five language subtypes weights are genetically correlated with Executive function behavioural cluster, but the highest genetic correlation is between Mix2 and subtypes S1,S3 and S5 (ρ_{gS1} = -0.5, ρ_{gS3}= -0.5 and ρ_{gS5}= 0.5). See supplementary material figure Sup 19 for a list of abbreviations used for the behavioural tests. c - Genetic correlation between working memory subtypes and in-scanner behavioural tests. Working memory subtypes are genetically correlated with all tasks either with reaction time or accuracy measures. Genetic correlation range from -0.8 to 0.7.

d - Genetic correlation between working memory subtypes and out-of-scanner behavioural measures: Genetic correlation estimate between working memory subtypes weights and out-of-scanner task performance shows the highest value for Executive function behavioural cluster followed by academic skills. 83

18 **Heritability of framewise displacement (FD) measure before and after scrubbing on the 7 tasks.** WM: working memory, SOC: social, REL: relational, MOT: motor, LAN: language, GAM: gambling and EMO: emotional. Heritability of FD was estimated before and after scrubbing to shows how motion reduction used in this paper is efficient in reducing the heritability estimate that is known to affect further estimate of heritability on brain phenotypes. 92

19 short caption 95

20 short caption 97

20 **Task activation subtypes:** Fives subtypes maps (S1 to S5) for each of the social, motor, emotion, gambling and relational tasks. Working memory and language subtypes are shown on the manuscript figure 14 99

20 **Genetic and phenotypic correlation for in-scanner behavioural performance:** Left column represents genetic correlation between in scanner behavioural performance and brain subtypes weights. Right column represents phenotypic correlation between in scanner behavioural performance and brain subtypes weights. 102

20 **Genetic and phenotypic correlation for in-scanner behavioural performance:** Left column represents genetic correlation between in scanner behavioural performance and brain subtypes weights. Right column represents phenotypic correlation between in scanner behavioural performance and brain subtypes weights. 105

21 **meta-subtypes generation and association with behavioural measures-a**From each subject, an un-thresholded statistical parametric map (SPM) is vectorized then stacked. **b**. Nuisance regression from stacked SPM (FD: Frame Displacement, Sys: Systolic blood pressure, Dias: Diastolic blood pressure, BMI: Body Mass Index). **c**. From the residual stack, a subject by subject correlation matrix is clustered into subgroups of activations maps (subtypes). **d**. Each subtype is represented spatially by a mean activation map of all subjects within the same cluster . **e**. Each subtype is represented on the subject level by its weight.

The subtypes weight is the correlation between a vectorized subject activation map and a vectorized subtype activation map. **f**. All subtypes weights from the 7 tasks are stacked into subjects by subtypes' weight matrix. **g**. A subtypes weight by subtypes weight genetic correlation matrix is clustered into subgroups of genetic meta-subtypes. **h**. Each meta-subtype is represented spatially by a mean map of all subtypes within the same cluster. **i**. Mean weights of genetic meta-subtypes are genetically, environmentally and phenotypically associated with behavioural measure. 113

22 short caption 117

23 **Genetic meta-subtypes association with behavioural measures - a - Phenotypic association between genetic meta-subtypes and behavioural measures:** Phenotypic correlation between genetic meta-subtypes (2,8) and clusters of behavioural measures. GenMetSub 2 and 8 shows a significant association with behavioural clusters related to self-regulation, executive function, positive effects and speed processing. **b - Genetic association between genetic meta-subtypes and behavioural measures:** Genetic correlation between genetic meta-subtypes (2,8) and clusters of behavioural measures. GenMetSub 2 and 8 shows a significant association with behavioural clusters related to self-regulation and executive functions. **c - Environmental association between genetic meta-subtypes and behavioural measures:** Environmental correlation between genetic meta-subtypes (2,8) and clusters of behavioural measures. GenMetSub 2 and 8 shows a no significant association with behavioural clusters. 118

24 **Heritability of framewise displacement (FD) measure before and after scrubbing on the 7 tasks.** WM: working memory, SOC: social, REL: relational, MOT: motor, LAN: language, GAM: gambling and EMO: emotional. Heritability of FD was estimated before and after scrubbing to shows how motion reduction used in this paper is efficient in reducing the heritability estimate that is known to affect further estimate of heritability on brain phenotypes. 125

25 short caption 127

26 short caption 129

27 **Spatial meta-subtypes maps** 129

Liste des sigles et des abréviations

HCP	Human Connectome Project
IRMf	Imagerie par résonance magnétique fonctionnelle
QC	Quality Control en anglais ou contrôle de qualité
FD	Frame Displacemtent

Remerciements

À Marji, Anana, Pixi et Camille

Chapitre 1

Introduction

Le comportement humain est le résultat de l'action conjointe de la nature et de la culture. La nature correspond au bagage génétique à la base de notre biologie, tandis que la culture fait référence aux influences environnementales (physiques et sociales) auxquelles les individus sont exposés depuis la conception. Comprendre comment l'action conjointe de la nature et de la culture façonne le comportement humain constitue l'un des principaux objectifs de l'anthropologie biologique. Cette discipline se base sur une approche évolutionnaire du comportement. Elle formule l'hypothèse générale que le comportement est le résultat de la sélection naturelle et sujet à évolution. Elle cherche ainsi à démêler la relation entre l'universalité (due à la nature) et la variabilité (due à la culture) de l'ensemble des comportements humains.

Une méthode permettant d'étudier cette relation entre universalité et variabilité est de mesurer les activations cérébrales associées au comportement, ainsi que les influences génétiques et environnementales sur ces dernières. L'imagerie par résonance magnétique fonctionnelle (IRMf) est une technique non invasive permettant l'exploration des activations cérébrales qui apparaissent liées – et qui sous-tendraient donc – le comportement. La génétique quantitative est une discipline qui permet de discerner les influences génétiques des influences environnementales sur le comportement. Ainsi, l'IRMf, combinée à la génétique quantitative [1], permet (1) de remonter à la source du comportement au plan cérébral et (2) de caractériser les influences génétiques et environnementales sur ses corrélats neurobiologiques. L'étude du comportement humain au moyen de l'imagerie cérébrale et de la génétique apparaît donc comme une voie particulièrement prometteuse pour l'anthropologie biologique moderne.

[1] La génétique quantitative est l'une des disciplines de la génétique qui étudie les mécanisme héréditaires des traits phénotypiques quantitatifs.

L'objectif principal de cette thèse est d'étudier la variabilité et l'universalité du comportement humain à travers les lentilles de l'IRMf en estimant l'influence génétique et environnementale sur divers patrons d'activation cérébrale en lien avec des tâches comportementales appartenant à différents domaines socio-cognitifs d'importance.

1.1. Le comportement humain : définitions

Le dictionnaire de psychologie d'Oxford (2009) définit le comportement comme :

> «*The physical activity of an organism, including overt bodily movements and internal glandular and other physiological processes, constituting the sum total of the organism's physical responses to its environment (Colman, 2009, p. 472)*»

Le Grand Dictionnaire de la Psychologie 1999 propose quant à lui la définition suivante :

> «*Le comportement est une réalité appréhendable sous la forme d'unité d'observation, les actes, dont la fréquence et les enchaînements sont susceptibles de se modifier ; il traduit en action l'image de la situation telle qu'elle est élaborée, avec ses outils propres, par l'être que l'on étudie : le comportement exprime une forme de représentation et de construction d'un monde particulier (Umwelt) (Bloch, 1999, p. 176)*»

Ces définitions renvoient à un élément important dans le rôle du comportement à savoir l'expression d'une dimension interne vers une dimension observable ou appréhendable. La dimension interne correspond à la représentation construite ou perçue du monde externe. Cette représentation du monde est propre à chaque espèce et connue sous la notion d'Umwelt (Bueno-Guerra, 2018; von Uexküll, 2010). Chez les espèces dotées d'un système nerveux, l'Umwelt est la façon dont l'organisme se représente le monde neurologiquement, tel que basé sur son spectre perceptuel (Merkwelt) et moteur (Wirkwelt) (von Uexküll, 2010). Chez les espèces sociales, l'Umwelt est enrichi par la sphère sociale (Kull, 2010), c'est-à-dire par l'ensemble des interactions entre individus. L'Umwelt chez notre espèce serait donc particulièrement complexe et enrichi par des capacités cérébrales propres à l'être humain, telles que l'abstraction, le langage et la fonction symbolique. Cette complexité inégalée de l'être humain rend l'interprétation du comportement observable difficile, de sorte qu'un simple clin d'œil, par exemple, peut s'expliquer d'une multitude de façons différentes (Geertz and Others, 1973). Il devient ainsi plus aisé d'étudier le comportement humain en faisant référence à l'état mental sous-tendant ses manifestations.

Dans cette thèse, le comportement humain est défini comme l'ensemble des activités psychiques et motrices qui influencent ou régissent les interactions de l'individu avec son environnement propre. Selon cette définition adaptée de Chapais et Pérusse 1988 sur l'origine du comportement humain, la pensée, les

émotions, les gestes, la parole, etc., sont des phénomènes comportementaux se manifestant à des moments particuliers de la séquence allant de l'activité neurobiologique de l'individu à ses interactions avec son environnement propre (Umwelt).

1.2. Neurosciences du comportement humain

Les progrès considérables en techniques et méthodes d'imagerie cérébrale accomplis au cours des dernières décennies ont permis de mieux comprendre les structures et le fonctionnement du cerveau en lien avec le comportement. Une technique d'imagerie en particulier a vu sa popularité augmenter dans le domaine des neurosciences comportementales : l'imagerie par résonance magnétique fonctionnelle (IRMf). Il s'agit d'une technique non invasive (ne comportant pas d'intrusion dans le cerveau d'appareillage ou de substances chimiques) qui présente actuellement le meilleur compromis entre résolutions spatiale (la précision de la localisation de l'activité cérébrale) et temporelle (le laps de temps qui s'écoule entre l'activation cérébrale et son enregistrement). L'IRMf peut détecter un signal localisé dans une région cérébrale de quelques millimètres avec un délai temporel de quelques secondes. La figure 1, tirée de Sejnowski et collègues 2014, donne une estimation des régions spatio-temporelles étudiées par les différentes méthodes de neuroimagerie fonctionnelle utilisées jusqu'en 2014. On remarque le vide que l'IRMf vient combler depuis quelques décennies en comparaison d'autres techniques existant avant l'arrivée de cette dernière méthode (figure 1, en bas à droite). Aujourd'hui, la majorité des recherches liant le comportement humain (sain ou pathologique) à ses corrélats biologiques observés dans le cerveau utilisent l'IRMf, malgré le coût élevé associé à ce type de cueillette des données.

1.2.1. IRMf

L'IRM est une technique qui permet de saisir à la fois l'anatomie (IRM structurelle) et le fonctionnement (IRM fonctionnelle ou IRMf) du cerveau. L'IRM structurelle fournit une image en trois dimensions (3D) de la structure anatomique du cerveau, alors que l'IRMf produit une série d'images en 3D reflétant l'intensité de l'activité cérébrale dans une région particulière pendant un certain laps de temps. L'IRMf se base sur l'association entre une tâche exécutée dans l'appareil d'imagerie cérébrale (scanner) et le changement de signal dans les régions cérébrales sollicitées. Les tâches effectuées dans le scanner peuvent être passives telles que la visualisation d'images, de mots, de vidéos, l'écoute de sons, etc., ou actives telles que la résolution de problèmes mathématiques, la mémorisation de mots, l'association d'éléments, la lecture, etc.

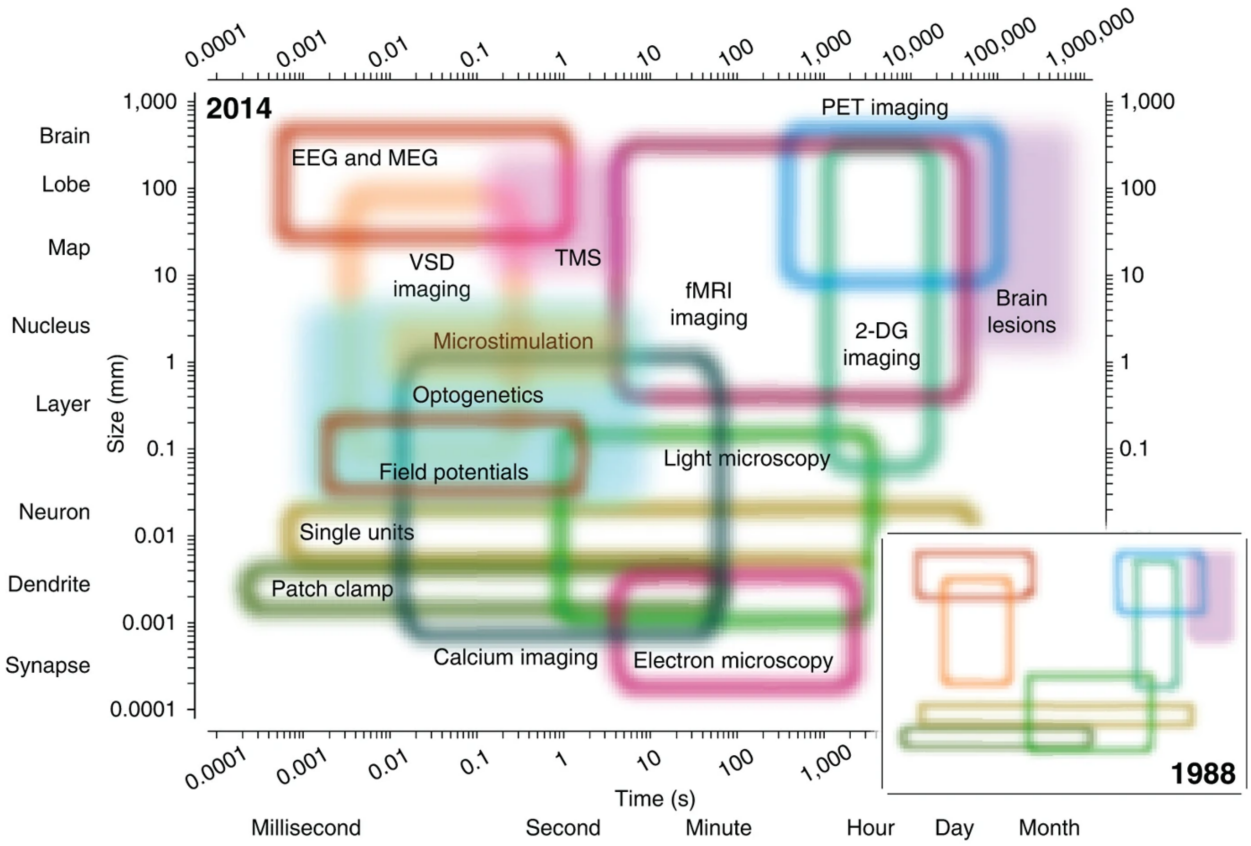


Fig. 1. Le domaine spatio-temporel en neurosciences et les principales méthodes disponibles pour l'étude du système nerveux en 2014 (tiré de Sejnowski et collègues 2014). Chaque région colorée représente le domaine spatio-temporel de chaque méthode d'imagerie fonctionnelle du cerveau jusqu'en 2014. Les régions vides illustrent l'utilisation de techniques de mesure, et les régions pleines de méthodes de perturbation (principalement invasives). En bas à droite, une représentation des méthodes disponibles en 1988 fait remarquer le vide que l'IRMf est venue combler en moins de deux décennies. **EEG** électroencéphalographie ; **MEG**, magnétoencéphalographie ; **TEP**, tomographie par émission de positrons ; **VSD**, colorant sensible au voltage ; **TMS**, stimulation magnétique transcrânienne ; **2-DG**, 2-désoxy-glucose.

L'IRMf évalue l'intensité de l'activité cérébrale de manière indirecte par la mesure des variations du signal BOLD (Blood-Oxygenation-Level-Dependent signal). L'activité neuronale dans une zone du cerveau provoque une augmentation de la consommation d'oxygène et du débit sanguin dans les vaisseaux qui irriguent cette région. L'augmentation du débit sanguin, chargé en oxyhémoglobine (hémoglobine porteuse d'oxygène) dans une région du cerveau, provoque une baisse relative de la concentration en désoxyhémoglobine (hémoglobine non chargée en oxygène) par rapport aux zones cérébrales non sollicitées (Habas, 2002). L'oxyhémoglobine est diamagnétique de par la liaison de son atome de fer avec des atomes d'oxygène. Inversement, la désoxyhémoglobine est paramagnétique, car

elle ne contient plus d'oxygène. Ce changement du rapport oxy/désoxyhémoglobine dans les vaisseaux est détecté en IRMf par une légère hausse (de 1 % à 5 %) du signal BOLD (dépendant du niveau d'oxygène sanguin) dans une zone sollicitée du cerveau (Gore, 2003). Au cours d'une session d'imagerie cérébrale, plusieurs images BOLD sont acquises tout au long de l'expérience. Chaque image représente un volume du cerveau de quelques milliers d'unités cubiques appelées voxels. Chaque voxel représente la moyenne du signal BOLD dans une région cérébrale de quelques millimètres cubes, reflétant l'intensité de l'activité neuronale dans cette zone(Logothetis and Wandell, 2004).

Dans cette thèse, j'utiliserai les données d'IRMf recueillies dans le cadre du consortium Human Connectome Project (voir chapitre 2) afin d'étudier certains corrélats neurofonctionnels du comportement humain.

1.3. Universalité et variabilité en IRMf du comportement

L'expansion du cerveau durant l'évolution de notre espèce a donné lieu à une augmentation de la taille, de la complexité et de la structure des circuits cérébraux en comparaison des autres primates. Cette expansion a principalement affecté le cortex cérébral, et plus particulièrement le cortex associatif (Fig 2A) (Fjell et al., 2013). Mueller et collègues 2013 ont démontré que cette zone associative du cerveau présenterait aussi une forte variabilité en termes de connectivité fonctionnelle au repos[2] comparativement aux régions primaires (Fig 2B). Ces auteurs ont aussi observé que la présence d'une variabilité entre individus d'une même espèce (variation inter-sujets) dans la mesure de connectivité fonctionnelle était fortement corrélée au degré d'expansion corticale durant l'évolution (Fig 2C), suggérant une origine évolutionnaire, donc d'ordre génétique, à la variabilité cérébrale fonctionnelle.

D'autre part, Il a été démontré que la région corticale associative présente le plus de plasticité (ou croissance) durant le développement postnatal (Buckner and Krienen, 2013; Fjell et al., 2013; Hill et al., 2010). Le cortex associatif reste immature pendant la gestation, et devient myélinisé plus tard dans le développement (Watson et al., 2006). Cette maturation prolongée expose le cerveau aux impacts de l'environnement pendant la période de développement. Le cerveau en développement est donc fortement marqué par les événements de la vie ainsi que les facteurs de stress (Lupien et al., 2009), ce qui suggère une origine environnementale à la variabilité cérébrale dans ces aires associatives. Il est donc important de départager les influences génétiques et environnementales à la base de la

[2]Cette technique consiste à scanner des sujets en IRMf en leur donnant comme directive de rester dans un état de repos (sans tâche à exécuter)

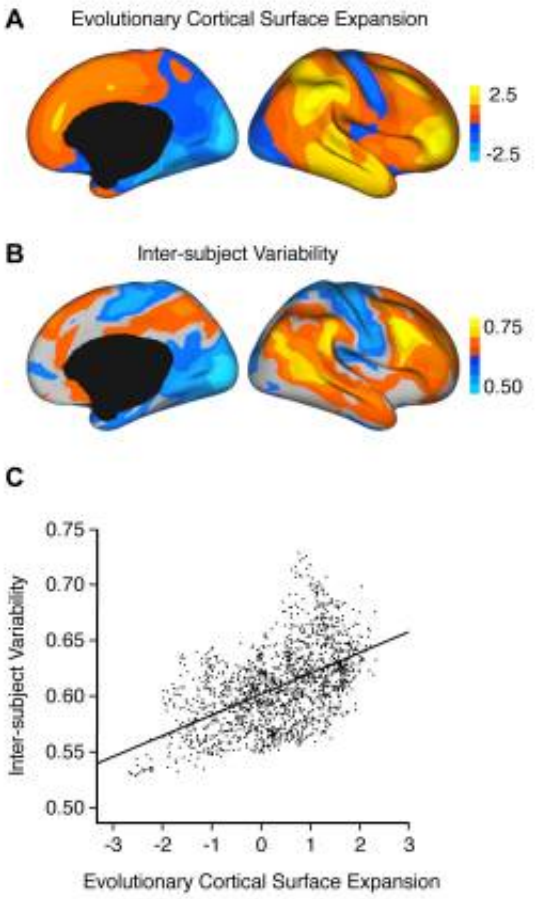


Fig. 2. Variabilité en connectivité fonctionnelle et expansion corticale évolutive. **A** - Expansion corticale entre un macaque et un humain. **B** - Variabilité fonctionnelle inter-sujets. **C** - Corrélation élevée (r = 0,52) entre l'expansion corticale et la variabilité fonctionnelle. (Mueller et al., 2013)

variabilité et de l'universalité fonctionnelles cérébrales.

Les études en IRMf du comportement (voir chapitre 2) font généralement appel à la moyenne des effets observés chez chaque sujet afin de produire des cartes d'effets de groupe (voir figure 5). Cette méthode repose sur deux prémisses : (1) la variance intra-individuelle (chez un même individu) des estimations de l'effet est uniforme dans un groupe de sujets (Penny and Holmes, 2007), et (2) on suppose que les estimations de l'effet de groupe suivent une distribution gaussienne, c'est-à-dire qu'il n'y a pas de valeurs atypiques (Chen et al., 2012). Au plan biologique, ces prémisses laissent supposer que chez les individus formant un groupe (p. ex. : des femmes adultes) les mêmes régions cérébrales s'activent uniformément lorsque les membres du groupe sont exposés à une même tâche (p.ex. : la lecture d'un texte) . Cette hypothèse a été mise à l'épreuve tant sur le plan méthodologique que neuroscientifique. Les irrégularités du scanner ou les artefacts

influençant la réponse BOLD peuvent conduire à la violation des prémisses susmentionnées (Mumford and Nichols, 2009; Penny and Holmes, 2007), alors que les différences réelles entre sujets dans les stratégies cognitives, l'attention portée à la tâche ou l'habituation à celle-ci, peuvent également introduire une diversité dans l'estimation des effets (Chen et al., 2012).

Dans une étude d'IRMf comportementale, Miller et collègues 2009; 2012 ont manipulé chez les sujets l'utilisation de stratégies verbales ou visuelles lors d'une tâche de mémorisation. Ils ont constaté que plus les sujets avaient tendance à utiliser des stratégies de mémorisation différentes (verbales contre visuelles), plus la différence était importante dans leurs patrons d'activation cérébrale liés à la tâche. Ces différences d'activations connexes à des stratégies cognitives ou à des capacités différentes ont été démontrées dans plusieurs domaines socio-cognitifs importants (Feredoes and Postle, 2007; McGonigle et al., 2000; Heun et al., 2000; Machielsen et al., 2000; Miller et al., 2002, 2009; Parasuraman and Jiang, 2012; Seghier et al., 2008; Seghier and Price, 2009). De telles différences d'activation cérébrale entre individus s'observent non seulement lors de processus cognitifs complexes, mais aussi dans le cas de tâches plus simples. Par exemple, Seghier et collègues 2016 ont constaté que le cortex moteur primaire était absent de la carte d'activation de groupe lors d'une tâche motrice, mais qu'il était visible dans près de 50 % des cartes d'activation individuelles. Dans leur ensemble, ces études montrent que les cartes d'activation de groupe ne représentent pas précisément les systèmes neuronaux engagés par les individus qui composent ces groupes (Lebreton and Palminteri, 2016). Cependant, ces mêmes études révèlent la présence d'une certaine concordance dans l'activation cérébrale entre individus. En conclusion, les activations neurobiologiques de divers sujets engagés dans une même tâche ne sont ni complètement différentes ni complètement identiques.

Le regroupement d'individus basé sur la similitude de leurs patrons d'activité cérébrale (sous-types neurobiologiques) a récemment gagné en popularité dans le domaine clinique, tels que par l'identification de sous-types d'atrophie corticale dans la maladie d'Alzheimer (Badhwar et al., 2019a) et de sous-types de connectivité fonctionnelle au repos chez des patients autistes (Easson et al., 2019; Urchs et al., 2020). Le clustering hiérarchique ainsi que l'analyse en composantes indépendantes sont parmi les principales méthodes statistiques ayant été utilisées avec succès pour le regroupement de sujets en fonction de la similarité de leurs cartes d'activation cérébrale (Kherif et al., 2003). Kashyap et collègues 2019 ont ainsi étudié les sous-types de connectivité fonctionnelle au repos chez 788 sujets du Human Connectome Project. Ils ont estimé la composante commune à travers plusieurs sessions d'imagerie au repos et ont classé les sujets en fonction de la similitude ou dissimilitude de leurs patrons de connectivité cérébrale. Deux sous-groupes ont émergé qui démontrent une association distincte avec des mesures comportementales liées, par exemple, à la

consommation de marijuana, de drogues illicites, d'alcool, de tabac, ainsi qu'avec le trait de personnalité antisociale. Une autre étude de Kirchhoff et Buckner 2006 a montré que diverses stratégies cognitives (visuelles et verbales) utilisées par les sujets lors d'une tâche de mémoire étaient corrélées avec des réseaux d'activation cérébrale distincts. Ces auteurs ont également constaté que l'activité dans les régions du cerveau associée à l'utilisation de ces stratégies était corrélée à la performance des sujets à une épreuve de mémoire. Ces études démontrent que s'il existe une hétérogénéité des fonctions cérébrales, celle-ci peut bel et bien être caractérisée en sous-types neurobiologiques par l'utilisation de méthodes d'exploration de données non supervisées, telles que le clustering hiérarchique.

Malgré la popularité croissante des méthodes de sous-typage d'activation cérébrale en IRMf au repos, seules quelques études ont exploré l'utilisation potentielle de ce sous-typage dans l'IRMf de tâches. Par conséquent, plusieurs questions persistent : les sous-types d'activation cérébrale peuvent-ils rendre compte à la fois de la variabilité et de l'universalité des manifestations neurobiologiques dans différents domaines socio-cognitifs ? Quel est le lien entre ces sous-types cérébraux et les différences génétiques et comportementales entre individus ? Que représentent ces sous-types au plan comportemental ?

Dans cette thèse, j'utiliserai la méthode de sous-typage basé sur le clustering hiérarchique des activations cérébrales dans sept domaines socio-cognitifs d'importance. J'estimerai l'influence de la génétique et de l'environnement sur la variabilité de ces sous-types cérébraux ainsi que leur lien avec les résultats à diverses tâches comportementales se rapportant aux domaines socio-cognitifs à l'étude.

1.4. Objectifs spécifiques

Le chapitre 3 porte sur la mise au point d'un protocole méthodologique original permettant de réduire le bruit lié à l'acquisition des données en IRMf, et d'augmenter par le fait même leur fiabilité, étape indispensable à l'utilisation de ces données dans les études neurocomportementales comme celle faisant l'objet de cette thèse.

Le chapitre 4 porte sur la mesure de corrélats neurofonctionnels des comportements à l'étude. Il devient alors possible d'explorer à la fois la variabilité et l'universalité du comportement par la méthode de regroupement des individus selon leurs cartes d'activation cérébrale. Dans ce même chapitre, je procède à l'estimation des influences génétiques et environnementales sous-tendant l'appartenance aux différents groupes de sous-types

cérébraux, ainsi que les corrélations entre ces sous-types et les mesures comportementales.

Enfin, au chapitre 5, je procède à l'estimation des bases génétiques qui lient entre eux les divers sous-types cérébraux, et leur association aux différents domaines socio-cognitifs à l'étude.

	Buts	Objectifs	Hypothèse	Méthode
Chapitre 3	**Réduire la variabilité liée à la qualité du prétraitement en IRMf**	Développer un protocole de QC simple et concis pour le recalage des images prétraité	Un protocole simple sera facilement déployable et aiderait les évaluateurs à obtenir un contrôle visuel plus fiable	Établir des lignes directrices claires et simples à appliquer sur la manière d'évaluer le recalage des images recalé sans avoir recours à des connaissances préalables de l'anatomie du cerveau
		Valider le protocole de QC entre expert et non-expert en évaluation des images recalé	le consensus des non-experts serait cohérent avec les évaluateurs experts, puisque le protocole ne requiert aucune connaissance de l'anatomie du cerveau	Comparer les évaluations des experts et non-experts sur le recalage des images en utilisant le même protocole.
Chapitre 4	**Démontrer que la viabilité des sous-types d'activation est un phénotype intermédiaire entre les gènes et le comportement**	Estimer la corrélation entre les sous types et le comportement (dans le scanner et en dehors du scanner)	Les score de performances de chaque tâche comportementale dans le scanner serait associé aux sous types d'Activation cérébrales dérivé de la même tâche, tandis que les mesure comportementale hors scanner seraient associées à des tâches lié par des domaine cognitifs connexes	Utiliser une méthode corrélationnelle qui élimine l'effet familial dans les estimation de l'association.
		Estimer l'influence génétique des sous-types cérébraux et du comportement	Une héritabilité modéré et élevé à la fois des sous types et des mesure comportementales	Utiliser un devis familial pour estimer l'héritabilité
		Estimer l'influence génétique commune à la fois sur les sous-types et le comportement	Une corrélation génétique modéré entre les sous types et les mesures comportementale	Utiliser les méthode d'estimation de pléiotropie
Chapitre 5	**Estimer l'influence génétique commune a travers different domaine cognitif**	Tester l'existence d'une influence génétique commune et modulaire entre les sous-types cérébraux de différent domaine cognitif	Les sous-types de différents domaines cognitifs présenteront une corrélation génétique élevée et se regrouperont en métal-sous-types modulaires.	Utiliser les méthode de clustering hiérarchique sur l'estimation de la pléiotropie
		Estimer la corrélation génétique, environnementale et phénotypique entre les méta-sous-types et les mesures comportementales	Une association significative avec le comportement	Utiliser les méthode d'estimation de pléiotropie et de corrélation phénotypique et environnementale

Tableau 1. Objectifs - Le tableau ci-dessous résume les objectifs spécifiques de chaque chapitre :

Chapitre 2

Méthodes et concepts dans l'analyse des phénotypes neurocomportementaux

Ce chapitre présente des aspects techniques utilisés pour l'analyse des données issues de l'IRMf. Chacun des articles qui suivent reprend certains de ces aspects en détail ; la lecture du présent chapitre n'est donc pas essentielle à la compréhension générale de la thèse.

2.1. La collecte de données

L'acquisition des images en IRMf est une étape importante et complexe qui requiert la mobilisation de ressources très coûteuses telles que la machinerie en IRM, le personnel qualifié et le recrutement des participants. Durant les années 1990 et jusqu'au début des années 2000, les études en IRMf comptaient peu de participants par étude. Une moyenne de 30 participants était le standard dans ce domaine. À cette époque, des données d'IRMf propres à chaque étude étaient collectées. Ces données étaient alors analysées, puis archivées. Ce modèle concervateur avait plusieurs désavantages. Par exemple, le petit nombre de sujets ne permettait généralement pas d'atteindre une puissance statistique suffisante (Button et al., 2013) mais aussi, il était difficile de répliquer les résultats d'autres équipes de recherche ou d'analyser les mêmes données avec des nouvelles méthodes et questions de recherche.

À partir des années 2010, les membres de la communauté spécialisée en neuroimagerie fonctionnelle prennent conscience de l'avantage du partage de données d'imagerie cérébrale afin d'accélérer les découvertes dans le domaine des neurosciences (Biswal et al., 2010). Des initiatives de partage voient le jour telles que OpenFMRI (aujourd'hui OpenNeuro) (Gorgolewski et al.), Nathan Kline Institute (NKI) (Nooner et al., 2012) ou International Neuroimaging Data-sharing Initiative (INDI) (Mennes et al., 2013). Ces initiatives partagent les données acquises en IRMf de diverses études ou centres avec l'accord des investigateurs principaux. Depuis, le nombre de publications utilisant les bases partagées a augmenté considérablement. Un exemple qui illustre bien cet effet est l'étude de Milham et collègues

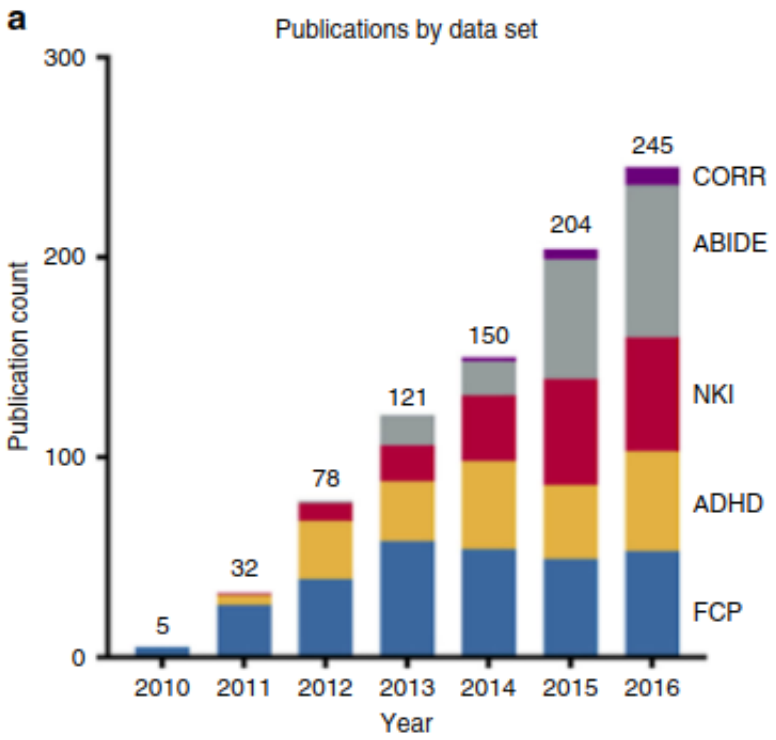


Fig. 1. Nombre de publications utilisant l'International Neuroimaging Data-sharing Initiative - INDI (tiré de Milham et collègues, 2018). Le nombre de publications scientifiques recensées entre 2010 et 2016 dans des revues à révision par les pairs dans lesquelles les différentes bases de données partagées de INDI ont été utilisées. **CORR**: Consortium for Reliability and Reproducibility, **ABIDE**: Autism Brain Imaging Data Exchange, **NKI**: Nathan Kline Institute-Rockland Sample, **ADHD**: Attention Deficit Hyperactivity Disorder Sample, et **FCP**: 1000 Functional Connectomes Project.

(Milham et al., 2018) sur l'impact du partage de données d'imagerie cérébrale sur la littérature scientifique. Les auteurs ont quantifié le nombre de publications dans des revues à révision par les pairs qui ont utilisé les données partagées par INDI. On observe une augmentation nette et quantifiable d'année en année du nombre de publications utilisant les bases de données de INDI (voir figure 1).

Cependant, il reste encore des défis envers lesquels faire face pour cette nouvelle ère de partage, à savoir : (1) les données sont acquises dans différents sites avec des appareils d'imagerie de différents modèles et spécificités (2) le contrôle de la qualité des images acquise n'est pas standardisé à travers la communauté et enfin (3) l'hétérogénéité des données, principalement les IRMf de tâches, rend une analyse groupée entre différentes bases de données difficilement validable (David et al., 2013). Une solution serait le partage des données de haute qualité par des consortiums en neuroimagerie qui auraient la mission de produire des données directement partageables à la communauté (Nichols et al., 2017).

La figure 2, tirée de l'article de Poldrack et Gorgolewski 2014, montre une estimation de la réutilisation des données par la communauté en fonction du coût par volume de données à partager. Ces auteurs estiment que plus les coûts de production et de partage des données sont élevés, meilleure est la qualité des données, ce qui augmenterait le potentiel de réutilisation par la communauté.

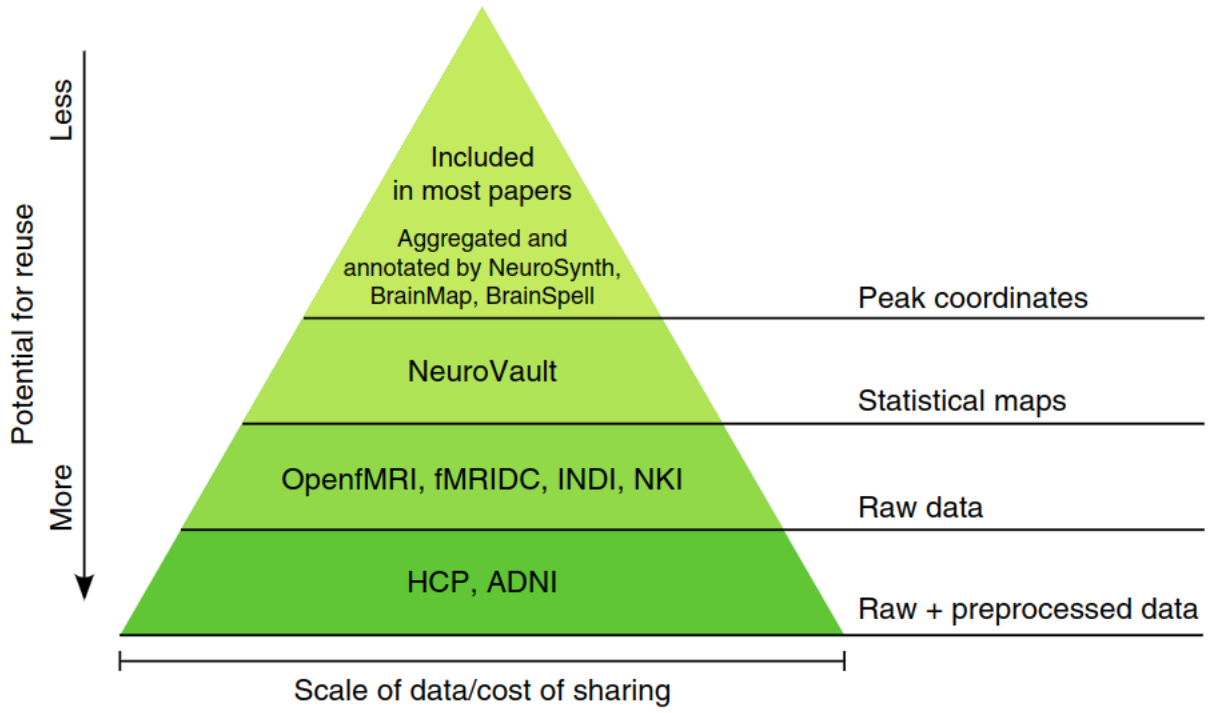


Fig. 2. Le potentiel de réutilisation de différents types de partage de données. Plus le coût et le nombre de données à partager sont élevés, plus le potentiel d'utilisation par la communauté est élevé. Tirée du texte de Poldrack et Gorgolewski 2014

Aujourd'hui, quelques grands consortiums et initiatives ont vu le jour. Ils ont pour but principal la mise à disposition immédiate des données acquises (ou prétraitées) en IRMf pour la communauté scientifique (Glasser et al., 2016; Sudlow et al., 2015) . Une initiative importante dans cette lignée est le Human Connectome Project (HCP) (Van Essen et al., 2013). Ce projet rassemble des images recueillies par IRM de haute résolution chez 1200 jeunes adultes et mises en libre accès. La richesse de cette base de données réside dans le fait que les individus sont soit des jumeaux monozygotes ou dizygotes ainsi que leurs fratries (les frères et sœurs). Tous les sujets scannés passent par une batterie de tests psychologiques et comportementaux dont les résultats sont aussi mis à la disposition de la communauté scientifique (Van Essen et al., 2013). En ce qui a trait à l'imagerie fonctionnelle, chaque sujet a exécuté sept tâches différentes lors de la prise d'images : (1) une tâche motrice (mouvement des doigts et des pieds), (2) une tâche émotionnelle (visualisation d'images

présentant différentes expressions faciales), (3) une tâche langagière (compréhension d'une histoire audio), (4) une tâche de mémoire (exercice de rappel de mémoire à court terme), (5) une tâche de prise de risque (jeux d'argent), (6) une tâche sociale (juger des interactions impliquant ou non une théorie de l'esprit) et (6) une tâche d'association (association entre des formes et symboles) (Barch et al., 2013).

Dans cette thèse, je vais utiliser les données en accès libre du HCP, et plus spécifiquement les images brutes captées par IRMf lors des sept tâches décrites plus haut, ainsi que des données issues des tests psychologiques et comportementaux.

2.2. L'analyse des données

Les images captées par IRMf représentent des données brutes qui nécessitent un certain nombre d'étapes de prétraitement afin d'être utilisées dans les analyses subséquentes . Ces étapes visent d'abord à réduire les divers artefacts qui compromettent l'interprétation des fluctuations du signal BOLD en IRMf, par exemple les artefacts physiologiques et de mouvement. Le second objectif majeur est d'aligner les données acquises à différents moments pour un même sujet, mais aussi d'établir une certaine correspondance entre les cerveaux de différents sujets, de telle sorte qu'une inférence quant au rôle d'une zone cérébrale donnée puisse être portée au niveau du groupe de sujets à l'étude. Certains logiciels de prétraitement de données en IRMf sont disponibles en accès libre, tels que FMRIPrep (Esteban et al., 2019) ou NIAK (Bellec et al., 2011).

Après le prétraitement, l'étape d'analyse permet de faire le lien entre les changements du signal BOLD dans tout le cerveau et les stimuli du devis expérimental. Afin d'établir ce lien, diverses méthodes statistiques et de calcul ont été proposées (Worsley, 1997), notamment le modèle linéaire général (GLM) largement utilisé pour l'IRMf de tâches en raison de son efficacité, sa simplicité et sa grande disponibilité (Friston et al., 1994). Une analyse de premier niveau (niveau individuel) qui consiste à faire convoluer le devis expérimental avec la réponse hémodynamique, permet de quantifier l'estimation du signal BOLD attendu pour toute stimulation dans le devis expérimental. Le but de ce processus est de générer un test statistique paramétrique pour chaque voxel du cerveau, ce qui permet de créer une carte statistique paramétrique du cerveau entier, pour chaque sujet. Enfin, pour faire des inférences au niveau du groupe, une deuxième analyse, aussi basée sur le GLM, permet de générer des cartes SPM pour le groupe entier. Cette approche GLM modélise les séries temporelles du signal de l'IRMf comme une combinaison linéaire de différentes composantes du signal, afin de tester si l'activité dans une région du cerveau est systématiquement associée

à une condition particulière d'intérêt dans le devis expérimental (Soares et al., 2016). Le GLM se calcule comme suit:

$$Y = X\beta + \varepsilon \tag{2.2.1}$$

Où Y correspond à la réponse BOLD enregistrée, X à la matrice du devis expérimental, β aux paramètres estimés et ε à l'erreur. Selon le type de devis expérimental, il est possible ensuite de soustraire les cartes SPM de différentes conditions au cours d'une même tâche. Cette technique permet d'isoler les processus cognitifs à l'étude. Par exemple, la figure 4 représente les régions cérébrales activées et désactivées suite à l'exécution d'une tâche de compréhension de langage moins une tâche de calcul mathématique (Barch et al., 2013).

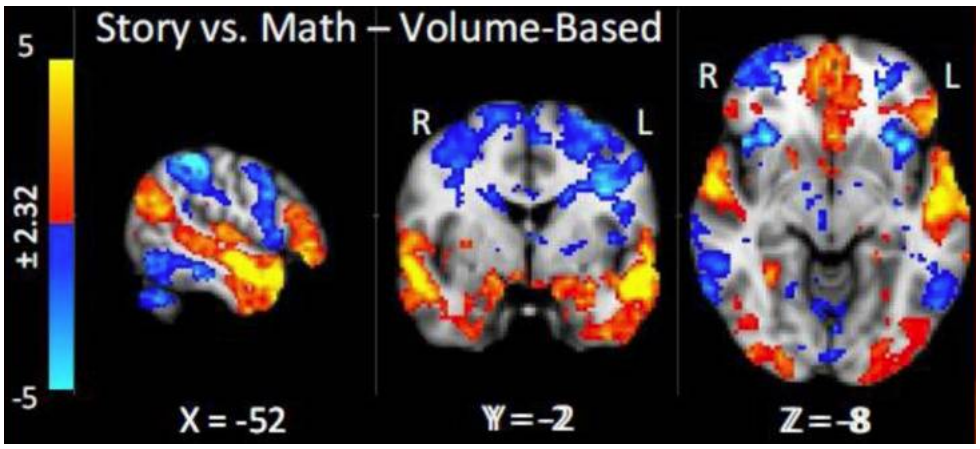


Fig. 3. Carte d'activation cérébrale d'analyse de groupe. Les zones cérébrales activées durant cette tâche représentent les régions connues pour le traitement de l'information auditive et du langage. Tiré de Barch et collègues 2013.

Enfin, il est important de noter que l'analyse (incluant le prétraitement) des grandes bases de données est exigeante en termes de ressources informatiques. Il est désormais impossible de mener une étude de l'envergure du HCP (1200 sujets) sur un ordinateur personnel. Pour pallier cette demande en ressources, certains pays ou universités mettent à la disposition de la communauté scientifique des installations informatiques dédiées au calcul de haute performance. Le Canada a mis des superordinateurs accessibles aux chercheurs des universités et collèges partout au Québec et au Canada. Ces installations permettent ainsi aux scientifiques d'accélérer la recherche dans plusieurs domaines, dont les neurosciences. (Baldwin, 2012)

Dans cette thèse je vais utiliser les données en accès libre du HCP, et plus spécifiquement les images brutes captées par IRMf lors des sept tâches décrites plus haut, ainsi que des données comportementales non issues de techniques

de neuroimagerie. J'analyserais les données d'activation cérébrale en GLM en utilisant les ressources de calcul Canada et Québec afin réduire le temps de calcul.

2.3. Les sources de variabilité liées à l'acquisition et le traitement des données en IRMf

L'IRMf en contraste BOLD est une technique largement utilisée en recherche malgré sa complexité en termes d'isolation du signal par rapport au bruit de mesure. Le signal est bruité par des éléments liés à la machine tels que le type d'appareil utilisé, la séquence d'acquisition, l'intensité du champ magnétique, mais aussi des éléments liés au bruit physiologique tels que la respiration, le battement cardiaque et les mouvements de la tête (Desjardins et al., 2001). Les facteurs liés à la machine et à l'acquisition des données provoquent des artefacts à la fois sur les images en T1 et BOLD, qui interfèrent ensuite avec les analyses des données. Ces artefacts sont visibles sur les images BOLD et se distinguent en trois sortes : (1) les distorsions spatiales (figure 6; G1), (2) les pertes de signal (figure 6; G2) et (3) les images fantômes (figure 6; G3). Les distorsions spatiales peuvent résulter de l'absence d'homogénéité du champ statique et apparaissent sur les images sous la forme de pixels étirés ou comprimés. La perte de signal est due à une inhomogénéité du champ dans les tissus cérébraux adjacents à des zones remplies d'air, comme les sinus. Enfin, les images fantômes se présentent sous la forme d'une deuxième image BOLD imprégnée dans l'image principale, et sont principalement dues à la technique d'acquisition.

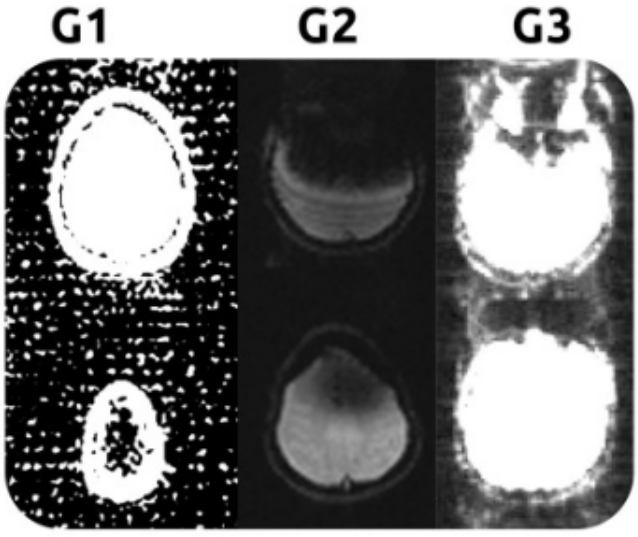


Fig. 4. Effet des artefacts sur les images BOLD - **G1**: Distorsion spatiale - **G2**: Perte de signal - **G3**: Images fantômes. Figure tirée de Soares 2016.

Un premier contrôle de la qualité (Quality Control ou QC en anglais) des images est généralement effectué lors de la séance d'acquisition dans le scanner. Cela permet de rescanner les sujets qui échouent le QC et de minimiser la perte de données. Un deuxième QC est exécuté après prétraitement des données. Le prétraitement comprend principalement le lissage des images, la correction des mouvements de la tête et la transformation spatiale des images. Ce QC a généralement pour but de vérifier si l'étape de transformation spatiale des images a bien été effectuée. Cette transformation spatiale, aussi appelée normalisation, consiste à aligner les images anatomiques (T1) et fonctionnelles (BOLD) de tous les individus dans un même espace de référence. Si les régions homologues du cerveau ne sont pas correctement alignées entre individus, la sensibilité est perdue, ce qui entraîne une augmentation du taux de faux négatifs (Lu et al., 2019). Cet alignement se fait principalement de manière automatique. Les algorithmes de normalisation des images sont susceptibles d'échouer (Dadar et al., 2018), et les résultats sont donc généralement examinés visuellement pour en garantir la qualité. La figure 7 montre des exemples de normalisation qui ont échoué de manière significative (Figure 7a) ou subtile (Figure 7b).

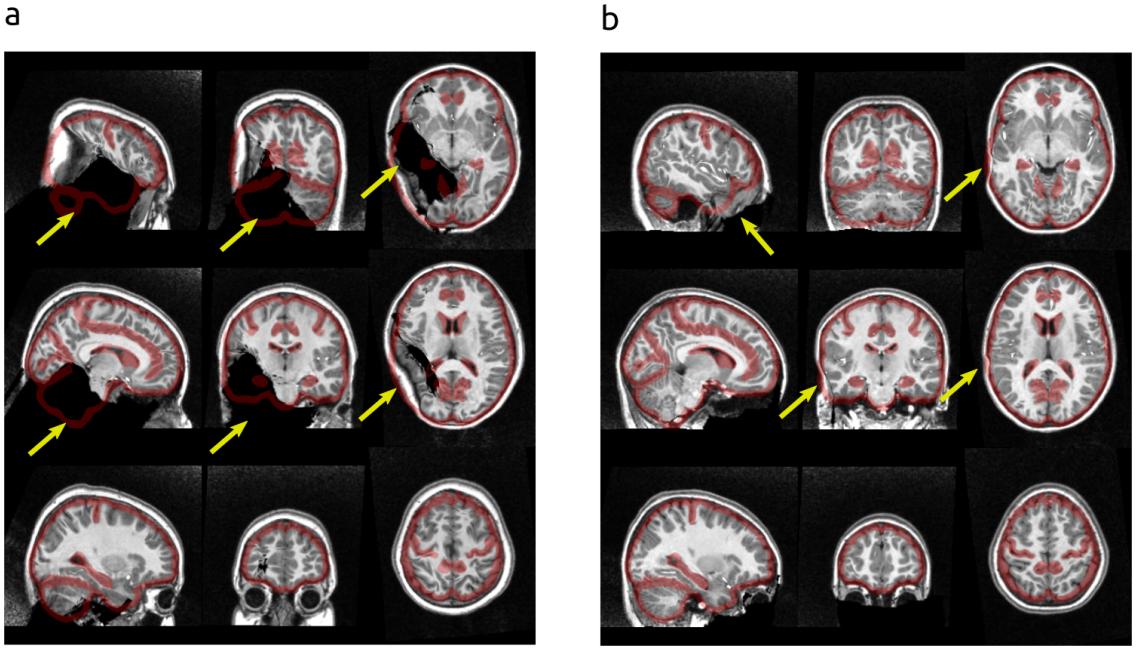


Fig. 5. Contrôle de qualité visuel de l'alignement des images T1 dans l'espace de référence. Les flèches jaunes indiquent les endroits où l'alignement a échoué. **a**: Sujet avec un alignement qui a échoué de manière claire. **b**: Sujet avec un alignement qui a échoué de manière subtile.

Il n'existe cependant pas de protocole standardisé et validé pour effectuer le QC visuel de l'alignement de la T1 dans l'espace de référence ni d'études robustes permettant d'estimer

la concordance entre évaluateurs lors d'une procédure de QC.

Dans cette thèse je vais réduire l'impact des erreurs de recalage des images T1 et BOLD en mettant au point un protocole de QC visuel simple à appliquer, tout en le validant avec des évaluateurs experts et non experts en QC de recalage des images.

2.4. Génétique quantitative du comportement humain

La génétique quantitative humaine est une discipline qui étudie les bases génétiques des phénotypes humains. Un phénotype inclut autant les traits morphologiques (stature, couleur des yeux, etc.) ou physiologiques (pression artérielle, fréquence cardiaque, contraction musculaire, etc.) que des caractères comportementaux, cognitifs et cérébraux (personnalité, quotient intellectuel, activité cérébrale, etc.) (Polderman et al., 2015). Le but principal de cette discipline est d'estimer les influences génétiques qui expliquent la variabilité interindividuelle des traits. Deux méthodes sont possibles pour estimer le rôle de la génétique dans la variabilité d'un trait : (1) l'étude d'association pangénomique (en anglais genome-wide association study, GWAS) qui consiste à comparer des séquences d'ADN entre individus ou populations et de les associer aux variations mesurées d'un trait morphologique, physiologique ou comportemental ; et (2) l'étude statistique de devis familiaux ou de jumeaux qui consiste à estimer le degré de ressemblance phénotypique entre apparentés attribuable à leur bagage génétique commun (Falconer et al., 1996). Dans ces deux méthodes, une métrique importante est extraite pour estimer l'apport génétique : l'héritabilité au sens large (H^2). Cette métrique représente le rapport de la variation génétique (Vg) relativement à la variation phénotypique totale (Vp) dans une population donnée (Falconer et al., 1996).

$$H^2 = \frac{Vg}{Vp} \tag{2.4.1}$$

Les principaux moteurs de la variation génétique sont les mutations et les recombinaisons génétiques (Griffiths et al., 2000). Chez les espèces sexuées, les recombinaisons génétiques se produisent seulement lors de la transmission de l'ADN maternel et paternel vers la progéniture. Cette recombinaison crée un nouvel arrangement des gènes chez la progéniture qui porte un mélange aléatoire de l'ADN parental.Quant aux mutations, il s'agit d'une altération de la séquence d'ADN qui survient par exemple au cours de sa réplication suite à l'exposition à certains rayonnements ou substances mutagènes. Ces mutations peuvent être bénéfiques pour l'organisme, délétères (nuisibles) ou neutres (n'ayant aucun effet sur son aptitude). Seules les mutations qui affectent les cellules sexuelles peuvent être transmises à la génération suivante. Dans le cas des mutations non délétères, les gènes acquièrent plusieurs variantes, appelées allèles, dans la population. Ces allèles d'un même gène codent

toujours le même trait (p. ex: la couleur des cheveux), mais elles diffèrent dans la manière dont le trait est exprimé (p. ex: cheveux bruns ou blonds). Les allèles peuvent coder des traits pour lesquels la variation est discrète (pouvant être séparée en catégories distinctes, comme la couleur des yeux), ou continue (comme la stature). Ce sont principalement les variations et la fréquence alléliques dans une population donnée qui sont visées par les études de génétique quantitative. Cette variation allélique est majoritairement responsable de la variation génétique dans une population.

En génétique quantitative les mesures et les concepts qui sont basés sur les groupes (ou populations) à l'étude sont exprimés en termes de fréquences alléliques ou de différence quantitative. La notion de valeur vient opérationnaliser ces concepts sur le plan individuel, de sorte que la mesure d'un phénotype sur un individu représente la valeur phénotypique (P) de cet individu (Falconer et al., 1996). Ainsi, les valeurs associées au génotype et à l'environnement sont, respectivement, la valeur génétique (G) et la déviation environnementale (E). On peut voir cette relation comme ceci : les gènes sont responsables d'une certaine valeur du trait à l'étude, et l'environnement (tout facteur non génétique) engendre la déviation par rapport à cette valeur. La valeur phénotypique (P) peut donc être exprimée comme suit :

$$P = G + E \tag{2.4.2}$$

Quant à la valeur G, elle est composée de la somme des valeurs des gènes à effet additif (A) sur la valeur phénotypique, des gènes à effet de dominance (D) et à effet d'épistasie ou d'interaction (I) (Neale et al., 1992). Ainsi G peut être exprimée comme suit :

$$G = A + D + I \tag{2.4.3}$$

En génétique quantitative, toutes ces valeurs (formule 2.4.2 et 2.4.3) individuelles sont exprimées dans la population en tant que la somme des déviations par rapport à la moyenne populationnelle, ce qui constitue la variance. Ainsi, la variance phénotypique (V_P) d'un trait représente la somme des variances des gènes à effets additifs (V_A), de dominance (V_D) et d'interaction (V_I), ainsi que la variance engendrée par l'environnement (V_E):

$$V_P = V_A + V_D + V_I + V_E \tag{2.4.4}$$

Étant donné que l'effet de V_I est généralement minime et que V_A correspond à la variance de la valeur reproductive dans une population (Falconer et al., 1996), la covariance attendue entre une paire d'apparentés biologiques y_1 et y_2 s'exprime comme suit :

$$cov(y_1,y_2) = rV_A + uV_D \tag{2.4.5}$$

r est le coefficient de V_A et u celui de V_D correspondant au degré d'apparentement entre individus. Le tableau 1 affiche les valeurs que prennent ces coefficients pour les apparentés de premier, second et troisième degrés. Par exemple, selon l'équation (4), la covariance attendue entre une fratrie (deux frères) issue de mêmes parents est égale à $\frac{1}{2}V_A + \frac{1}{4}V_D$. Il est donc possible de déduire les valeurs de V_A et V_D si l'on connaît les liens de parentés entre individus. Cette approche est connue sous le nom de devis familial. Elle se distingue du devis de jumeaux par l'inclusion des apparentés de différents degrés. Le devis de jumeau se base seulement sur les jumeaux monozygotes et dizygotes.

Relationship		*Coefficient* r (of V_A)	*Coefficient* u (of V_D)
MZ twins		1	1
First-degree	Offspring: parent	$\frac{1}{2}$	0
	Full sib	$\frac{1}{2}$	$\frac{1}{4}$
Second-degree	Half sib Offspring: grandparent Uncle (aunt): nephew (niece)	$\frac{1}{4}$	0
	Double first cousin	$\frac{1}{4}$	$\frac{1}{16}$
Third-degree	Offspring: great-grandparent Single first cousin	$\frac{1}{8}$	0

Tableau 1. Coefficients de la variance dans la covariance des apparentés (Tiré du livre de Falconer et Mackay 1996)

En génétique quantitative les gènes à effet additifs sont considérés comme ceux qui contribuent majoritairement à la variation phénotypique. La métrique principale qui rend compte de cet apport à la variation globale est l'héritabilité au sens restreint (h^2) dont l'équation est donnée par :

$$h^2 = \frac{V_A}{V_P} \tag{2.4.6}$$

L'héritabilité au sens restreint représente donc la proportion de la variance phénotypique qui est attribuable à des différences de valeur reproductive, ce qui reflète le degré d'influence génétique sur la ressemblance phénotypique d'un trait quelconque chez divers apparentés (Falconer et al., 1996). L'héritabilité varie entre 0 et 1 de sorte qu'une héritabilité proche de 0 indique que la variabilité d'un trait dans une population est majoritairement attribuable à une variabilité due aux facteurs environnementaux, et vice versa. Ceci dit, il serait erroné d'interpréter la valeur de l'héritabilité comme le degré d'influence génétique sur un phénotype. Par exemple, l'héritabilité du nombre de doigts d'une main humaine est presque nulle

alors que le trait en question est sous forte influence génétique (D. Pérusse, communication personnelle). En effet, toute la variabilité phénotypique observée dans une population est imputable à la variation environnementale telle que les accidents, infections ou amputation chirurgicale. La mesure de l'héritabilité restreinte a été largement utilisée dans les études de jumeaux et les devis de famille. Une vaste méta-analyse, menée par Polderman et collègues (Polderman et al., 2015), a regroupé les résultats de 2 748 publications (14 558 903 sujets) portant sur 17 804 traits incluant des caractères psychologiques, physiologiques, sociaux et démographiques. L'héritabilité moyenne, tous traits confondus, est estimée à 0,49, c'est-à-dire près de la moitié de la variation phénotypique. Pour la majorité des caractères (69%), la ressemblance entre paires de jumeaux monozygotes seraient seulement dues à l'influence de gènes à effets additifs.

Une fois l'héritabilité restreinte estimée, une autre mesure importante peut être déduite qui est la corrélation génétique (ρ_g), aussi connue sous le nom de pléiotropie. Cette mesure nous renseigne sur le lien génétique qui pourrait exister entre deux ou plusieurs phénotypes. Des gènes sont dits pléiotropiques lorsqu'ils ont la propriété d'affecter plusieurs caractères à la fois (p. ex. : le poids et la taille). Un exemple de pléiotropie documenté au niveau moléculaire chez l'être humain est la phénylcétonurie. Cette maladie est causée par une déficience de production d'une enzyme intervenant dans la cascade de transcription et de traduction de l'ADN en protéines. Une mutation ponctuelle dans le gène qui code cette enzyme entraîne de multiples phénotypes associés, dont une déficience intellectuelle progressive, l'eczéma et des défauts de pigmentation qui rendent la peau des malades plus claire (Paul, 2000). En génétique quantitative il est possible d'estimer la relation pléiotropique entre deux traits phénotypiquement corrélés en quantifiant le degré de chevauchement des effets moyens qui influencent ces traits (Boomsma, 1996; Almasy et al., 1997).

Dans cette thèse je vais estimer l'héritabilité restreinte de certains comportements humains ainsi que leurs corrélats neuronaux en utilisant un devis familial. J'estimerai également la pléiotropie entre mesures comportementales et neuronales.

First Article.

A standardized protocol for efficient and reliable quality control of brain registration in functional MRI studies

by

Yassine Benhajali[1], AmanPreet Badhwar[2], Helen Spiers[3], Jonathan Armoza[4], Thomas Ong[5], Daniel Pérusse[6], and Pierre Bellec[7]

([1]) Département d'anthropologie, Université de Montréal, Montréal, Québec, Canada.

([2]) Centre de recherche de l'Institut universitaire de gériatrie de Montréal (CRIUGM), Montréal, Québec, Canada.

([3]) The Zooniverse, Department of Physics, University of Oxford, Oxford, UK.

([4]) English Department, New York University, New York, USA.

([5]) Département de psychologie, Université de Montréal, Montréal, Québec, Canada.

([6]) Département d'anthropologie, Université de Montréal, Montréal, Québec, Canada.

([7]) Département de psychologie, Université de Montréal, Montréal, Québec, Canada.

This article was submitted in Frontiers.

Published in Frontiers the 28th of February 2020.

Abstract. Automatic alignment of brain anatomy in a standard space is a key step when processing magnetic resonance imaging for group analyses. Such brain registration is prone to failure, and the results are therefore typically reviewed visually to ensure quality. There is however no standard, validated protocol available to perform this visual quality control.

We propose here a standardized QC protocol for brain registration, with minimal training overhead and no required knowledge of brain anatomy. We validated the reliability of three-level QC ratings (OK, Maybe, Fail) across different raters. Nine experts each rated N=100 validation images, and reached moderate to good agreement (Kappa from 0.4 to 0.68, average of 0.54±0.08), with the highest agreement for "Fail" images (Dice from 0.67 to 0.93, average of 0.8±0.06). We then recruited volunteers through the Zooniverse crowdsourcing platform, and extracted a consensus panel rating for both the Zooniverse raters (N=41) and the expert raters.

The agreement between expert and Zooniverse panels was high (kappa=0.76). Overall, our protocol achieved good reliability when performing a two-level assessment (Fail vs OK/Maybe) by an individual rater, or aggregating multiple three-level ratings (OK, Maybe, Fail) from a panel of experts (3 minimum) or non-experts (15 minimum).

Our brain registration QC protocol will help standardize QC practices across laboratories, improve the consistency of reporting of QC in publications, and will open the way for QC assessment of large datasets which could be used to train automated QC systems.

Keywords: Task-fMRI, Subtypes, Heritability, Genetic Correlation, Behavioural measures.

1. Introduction

Aligning individual anatomy across brains is a key step in the processing of structural magnetic resonance imaging (MRI) for functional MRI (fMRI) studies. This brain registration process allows for comparison of local brain measures and statistics across subjects. A visual quality control (QC) of brain registration is crucial to minimize incorrect data in downstream analyses of fMRI studies. However, no standardized, validated protocol has yet been developed to streamline this QC. Here, we present a standardized procedure for visual QC of brain registration and describe the reliability of QC ratings from both expert raters and a large panel of non-experts recruited through an online citizen science platform (www.zooniverse.org).

1.1. Brain registration

MRI is a non-invasive technique that can be applied to study brain structure (sMRI) and function (fMRI). Multiple steps are required to transform raw MRI data to processed images ready for downstream statistical analyses. One critical preprocessing step is brain

registration; this involves aligning 3D brain images to a standard stereotaxic space, such as the MNI152/ICBM2009c template (Fonov et al., 2009). State-of-the-art registration procedures use non-linear optimization algorithms such as ANIMAL (Collins and Evans, 1997), DARTEL (Ashburner, 2007) or ANTS (Avants et al., 2009). Dadar and colleagues (Dadar et al., 2018) compared 5 publicly available, widely used brain registration algorithms in medical image analysis and found a failure rate of 16.8% ± 3.13% on their benchmarks. This lack of robustness is mainly due to differences in image quality, shape and cortical topology between individual brains. A visual QC of registered brain images is thus required to ensure good data quality for subsequent analyses.

1.2. Visual QC

The specific focus of the visual QC for sMRI registration depends on the intended use of the data. Voxel-based analysis of brain morphology typically calls for a highly accurate registration, as this step can impact brain tissue segmentation. In contrast, fMRI studies usually rely on larger voxel size and spatial blurring, and are less likely to be affected by small registration errors. To our knowledge, as of yet, there are no standardized criteria for tolerable errors in sMRI registration for fMRI processing pipelines. Many fMRI analytical software packages present users with images to assess the quality of T1 image registration. In one of the most recent packages developed by the community, fMRIprep (Esteban et al., 2019), the registered T1 image is presented across 21 brain slices, along with images for three other processing steps (skull stripping, tissue segmentation, and surface reconstruction), yielding a total of 84 brain slices for visual inspection. Established processing tools like FMRIB Software Library (Jenkinson et al., 2012) or the Statistical Parametric Mapping Matlab package (Wellcome Centre for Human Neuroimaging) also present users with reports that often include more than ten brain slices for visual inspection for each subject. This makes visual inspection tedious and time-consuming. Critically, none of these packages offer guidelines on how to assess the quality of structural brain registration for fMRI studies. Without such guidelines and with a large number of images to review, QC is likely to vary significantly across raters.

1.3. Inter-rater agreement

QC studies of preprocessed images rarely report inter-rater reliability, and no such study examined brain registration to our knowledge. Pizarro and colleagues (Pizarro et al., 2016) applied a support vector machine algorithm on visually rated (N=1457 usable/unusable) sMRI data from 5 to 9 investigators who rated the same 630 images, but did not report agreement metrics. White and colleagues (White et al., 2018) compared automated QC metrics and manual QC from 6,662 sMRI data from 4 different cohorts/sites, merging visual

inspection across sites, raters, protocols and scan quality but without presenting agreement statistics. Studies that do report inter-rater agreement mostly focus on issues related to raw MRI images (e.g., signal-to-noise ratio or susceptibility artifacts), head motion (e.g., ghosting or blurring), brain extraction, and tissue segmentation. Inter-rater agreement in these studies is found to vary considerably. For example, Backhausen and colleagues (Backhausen et al., 2016) reported high agreement for two trained raters who visually inspected the same 88 sMRI,achieving an intra-class correlation of 0.931 for two categories of quality (pass-fail) on issues related to MRI acquisition and head motion. Esteban and colleagues (Esteban et al., 2017) reported a Kappa of 0.39 between two raters for three quality categories (Exclude/Doubtful/Accept) on 100 images when ratings were based on the quality of the MRI acquisition, head motion, brain extraction and tissue segmentation. Table 2 shows recent (2010 onwards) studies reporting inter-rater agreements on visual QC of sMRI for a variety of issues. Only one study, Fonov and colleagues (Fonov et al., 2018), included brain registration for visual QC assessment. These authors reported a test-retest Dice similarity of 0.96% from one expert rater who evaluated as pass or fail 1000 images twice, but no inter-rater reliability estimate. Variability in reliability across studies may be due to two types of factors: user- and protocol-related factors. Protocol-related factors (e.g., clarity, levels of rating or training set) can be addressed by multiple iteration and refinement of the protocol. Factors related to the rater (e.g., level of expertise, fatigue, motivation, etc.) are more difficult to constrain or control. One solution to circumvent individual rater variability is to aggregate multiple ratings from a large pool of raters.

1.4. Crowdsourced QC

Crowdsourcing can be used to achieve multiple QC ratings on large collections of images rapidly. Crowdsourcing, as first defined by Howe in 2016, is "*... the act of taking a job traditionally performed by a designated agent (usually an employee) and outsourcing it to an undefined, generally large group of people in the form of an open call*" (Howe, 2006). Crowdsourcing can be used in citizen science research projects where a large number of non-specialists take part in the scientific workflow to help researchers (Franzoni and Sauermann, 2014; Simpson et al., 2014). Crowdsourcing labor-intensive tasks across hundreds or thousands of individuals has proven to be effective in a number of citizen science research projects, such as modeling complex protein structures (Khatib et al., 2011), mapping the neural circuitry of the mammalian retina (Kim et al., 2014), and discovering new astronomical objects (Lintott et al., 2009; Cardamone et al., 2009).

In brain imaging, recent work by Keshavan and colleagues (Keshavan et al., 2018b) showed the advantages of using citizen science to rate brain images for issues related to

Study	QC agreement details		Reported visual inspection issue relate to: MRI Aquisition	Head motion	Brain extraction	Tissue segmentation	Brain registration
Backhausen et al. 2016	Nb. Images Nb. Raters Rating scale QC Manual Agreement	88 2 Include/Exclude http://bit.ly/Backh2016 ICC =0.93	Image sharpness, ringing, Contrast to noise ratio (subcortical structures and gray/white matter) and susceptibility artifacts	Ghosting or blurring	N.R	N.R	N.R
Esteban et al. 2017	Nb. Images Nb. Raters Rating scale QC Manual Agreement	100 2 Exclude/Doubtful/Accept N.R Cohen's Kappa =0.39	signal-to-noise ratio, Image contrast and Ringing	Head motion artefacts	Gray/white matter and the pial delineation	Graw-White matter segmentation	N.R
Rosen et al. 2018	Nb. Images Nb. Raters Rating scale QC Manual Agreement	Phase1=100, Phase2=100 Phase1=2, Phase2=3 0/1/2 N.R Phase1=100%, Phase2=85%	N.R	N.R	N.R	N.R	N.R
Fonov et al. 2018 (preprint)	Nb. Images Nb. Raters Rating scale QC Manual Agreement	9693 (1000 rated twice) 1 Accept/Fail Dadar et al. 2018 paper intra-rater Dice similarity =0.96	Effect of noise and image intensity non-uniformity	N.R	N.R	N.R	Incorrect estimates of, translation, scaling in all directions and rotation.
Klapwijk et al. 2019	Nb. Images Nb. Raters Rating scale QC Manual Agreement	80 5 Excellent/Good/Doubtful/Failed http://bit.ly/Klapw2019 Reliability =0.53	N.R	Ringing	Division between grey/white matter and pial surface	Grey-White matter segmentation	N.R

Table 2. Reported agreement in visual inspection of sMRI data on QC studies - QC studies since 2010 that uses sMRI and reported their inter/intra-raters agreement (Dadar et al., 2018; Esteban et al., 2017; Fonov et al., 2018; Klapwijk et al., 2019; Pizarro et al., 2016; Rosen et al., 2018)

head motion and scanner artifacts. These authors were able to gather 80,000 ratings on slices drawn from 722 brains using a simple web interface. A deep learning algorithm was then trained to predict data quality, based on the gathered rating from citizen science. The deep learning network performed as well as the specialized algorithm MRIQC (Esteban et al., 2017) for quality control of T1-weighted images. QC of large open access databases like HCP (Glasser et al., 2016), UKbiobank (Alfaro-Almagro et al., 2018) or ABCD (Casey et al., 2018) is challenging and time-consuming task if done manually. Using crowdsourced rating could be a key element to rate huge databases and possibly use these ratings to efficiently train a machine learning models to perform QC.

Here we propose a novel, standardized visual QC protocol for the registration of T1 images by non-experts. We formally assessed protocol reliability, first with "expert" raters familiar with visual inspection of brain registration, and second with a large pool of

"non-expert" raters with no specific background in brain imaging. These citizen scientists contributed via the world's largest online citizen science platform, called Zooniverse (Simpson et al., 2014). Zooniverse enabled the enrollment of more than 2000 volunteers from around the globe, thus enabling the evaluation of consensus between non-expert raters on a large scale. Specific aims and hypotheses of the study were as follows:

- To establish a QC procedure for MRI brain registration that does not require extensive training or prior knowledge of brain anatomy. Our hypothesis was that such a procedure would help raters achieve more reliable visual QC.
- To quantify the agreement between a consensus panel composed of non-expert raters and that of experts. Our hypothesis was that the consensus of non-experts would be consistent with experts' assessments, since the protocol requires no knowledge of brain anatomy.

2. Method

2.1. QC protocol building

The QC protocol was developed iteratively over the past 5 years, with several rounds of feedback from users. Initially, the protocol was used internally in our laboratory (Yassine and Pierre, 2016), and required a visual comparison of T1 slices against a template using the Minctool register (Janke and Fonov, 2012). Although the protocol achieved good consistency of ratings between two expert users (Kappa = 0.72), it was time consuming and hard to teach. We then switched from an interactive brain viewer to a static mosaic comprised of 9 different slices (3 axial, 3 sagittal, 3 coronal, see fig 1-b), and we highlighted anatomical landmarks using a precomputed mask. These landmarks were selected because we expected all of them to align well in the case of a successful registration, and the precomputed mask served as an objective measure to decide on the severity of a misalignment. We established guidelines on how to rate a registered image on a three-level scale ("OK", "Maybe" or "Fail") using these landmarks. The new protocol limited the need for extensive training for new users and potentially reduced the subjectivity of decision, notably for edge cases. The following sections describe the details and the validation of the final protocol (brain slices, landmarks and rating guidelines).

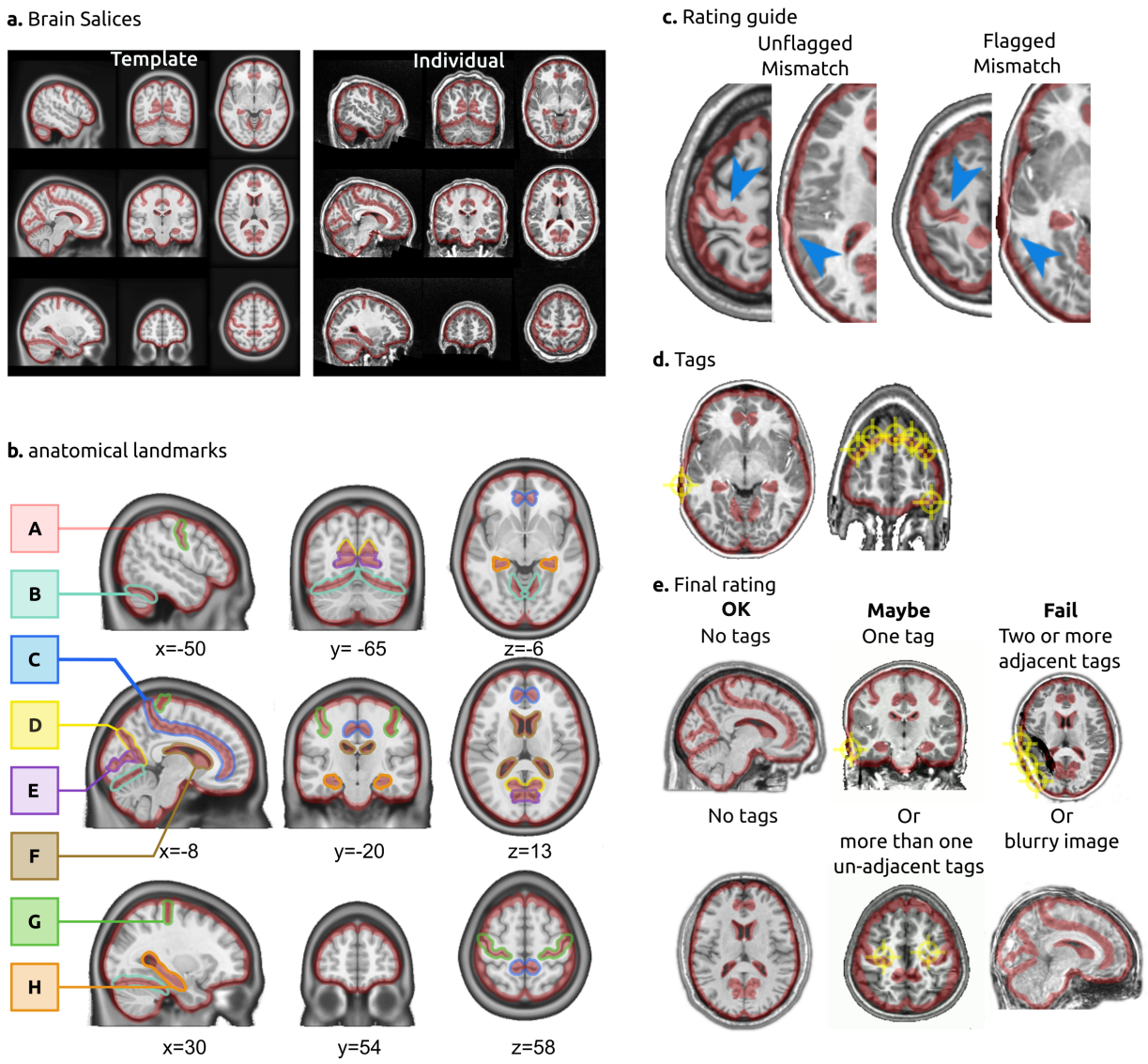


Fig. 6. QC protocol for brain registration- a: brain slices - he rater is presented with two sets of brain slices (3 axial, 3 sagittal and 3 coronal), one of them showing the template in stereotactic space and the other showing an individual T1 brain after registration. In the interface, the two images are superimposed and the rater can flip between them to visually assess the registration. **b: anatomical landmarks** - The landmarks for QC included: the outline of the brain (A), tentorium cerebelli (B), cingulate sulcu s(C), parieto-cingulate sulcus occipital fissure (D), calcarine fissure (E), the lateral ventricles (F), central sulcus (G) and the hippocampal formation (H) bilaterally. The landmarks were outlined in stereotaxic space. **c: rating guidelines** - The boundaries of red landmarks act as "confidence interval" for registration: an area is tagged as a misregistration only if the target structure falls outside the boundaries. **d: tags** - Raters put tags on each misregistered brain structure. **e: final rating** - A final decision is reached on the quality of registration: an image with no tags is rated OK, one or more nonadjacent tags are rated Maybe, two or more adjacent tags are rated Fail. An image that is excessively blurry is also rated Fail.

2.2. Brain slices

A mosaic view of 9 brain slices was extracted from each registered brain. The x, y and z coordinates, corresponding to axial, coronal and sagittal views, were as follows:

x (sagital)	***y (coronal)***	***z (axial)***
-50	-65	-6
-8	-20	13
30	54	58

Two images were generated: one using the individual T1 image of a subject, after brain registration, and one using the MNI2009c MRI T1 template averaged from 152 adults after iterative non-linear registration (Fonov et al., 2011b), see Figure 6a.

2.3. Anatomical landmarks

Notable anatomical landmarks included the central sulcus, cingulate sulcus, parieto-occipital fissure, calcarine fissure, tentorium cerebelli, lateral ventricles, bilateral hippocampal formation and the outline of the brain (see Figure 6b). To highlight these landmarks, we hand-drew a red transparent outline inside the brain with the MRIcron drawing tool (Rorden, 2014) using the MNI 2009 gray matter atlas as a reference. For the outline of the brain, a mask was generated using a 4-mm dilation of the template brain mask (as available in the MNI 2009c release, from which we subtracted a 4-mm erosion of the brain mask. This process resulted in a roughly 8-mm thick mask centered on the outline of the brain in template space. The landmark boundaries served as the "confidence interval" of acceptable registration. The width of this confidence interval was somewhat arbitrary, but critically helped raters to consistently assess what amount of misregistration was acceptable. The scripts to generate the mosaic brain images with highlighted landmarks have been made available in the GitHub repository (Brainmatch).

2.4. Rating guidelines

We instructed raters to focus on the brain structures within the red anatomical landmarks, comparing the individual brain, after registration, with the MNI 2009c template. The two images were presented superimposed with each other, and raters were able to flip manually or automatically between the individual and the template brain. For a given anatomical landmark, raters were asked to tag any part of the brain structure that fell outside of the anatomical landmark for the individual brain. The template acted as a reference for what the structure looked like, and where it was supposed to be. Figure 1c provides examples of acceptable and unacceptable registration of brain structures within the landmarks.

Raters were instructed to click all misregistered brain structures, which resulted in a series of tag spheres with 4 mm radius (Figure 6d). After an image was fully tagged, the overall registration quality was evaluated by the rater as follows:

- "OK" if no tag was reported,
- "Maybe" if one or several regions were tagged, yet no tag spheres overlapped (less than 8 mm apart),
- "Failed" if two tag spheres overlapped, meaning that an extensive brain area (>8 mm) was misregistered. Alternatively, a "Failed" rating was also issued if the entire image was of poor quality due to motion or a ringing artifact (Figure 6e)

2.5. Zooniverse platform

We used the online citizen science platform Zooniverse (Simpson et al., 2014) as an interface to perform the validation of our QC protocol[1]. Zooniverse offers a web-based infrastructure for researchers to build citizen science projects that require a human visual inspection and possibly recruit a large number of zooniverse volunteers, who are not familiar with neuroimaging and have no formal requirements to participate (Franzoni and Sauermann, 2014; Simpson et al., 2014). Our project, called "Brain match" was developed with the support of the Zooniverse team, to ensure compliance with Zooniverse policies and appropriate task design for an online audience[2], and the project was also approved by our institutional review board. Note that the raters were considered part of the research team, and not participants of the research project, and thus they were not required to sign an informed consent form. The project underwent a "beta review" phase on zooniverse, where we collected feedback on the clarity and difficulty level of the task. Rating was performed by Zooniverse raters and expert raters. All ratings were performed on the zooniverse platform through the Brain Match dashboard . The rating workflow was the same for the 2 types of raters. Note that individuals participating in Zooniverse choose to voluntarily dedicate some of their time to science and thus do not constitute a representative sample of the general population.

2.6. Brain images validation and training sets

We used a combination of two publicly available datasets, COBRE (Mayer et al., 2013) and ADHD-200 (Bellec et al., 2017), for both the beta and the full launch of the project. These datasets have been made available after anonymization by a consortium of research teams, each of which received ethics approval at their local institutional review board, as well as informed consent from all participants. Each individual sMRI scan was first corrected for intensity non-uniformities (Sled et al., 1998) and the brain extracted

[1]https://www.zooniverse.org/projects/simexp/brain-match
[2]https://www.zooniverse.org/lab-policies

using a region growing algorithm (Park and Lee, 2009). Individual scans were then linearly registered (9 parameters) with the T1 MNI symmetric template (Fonov et al., 2011b). The sMRI scans were again corrected for intensity non-uniformities in stereotaxic space, this time restricted to the template brain mask. An individual brain mask was extracted a second time on this improved image (Park and Lee, 2009) and combined with template segmentation priors. An iterative non-linear registration was estimated between the linearly registered sMRI and the template space, restricted to the brain mask (Collins et al., 1994). The processed data were finally converted into mosaics and merged with a mask of anatomical landmarks using in-house scripts. Two expert raters (PB,YB) rated each 954 preprocessed images in ADHD-200, achieving a kappa of 0.72 (substantial agreement) from a random subset of 260 images. The COBRE dataset was rated by YB only.

On Zooniverse, raters were first invited to read a tutorial (Supplementary material Figure S2) explaining the protocol, and then completed a QC training session, featuring 15 selected images (5 rated OK, 5 rated Maybe and 5 rated Fail, as rated by YB). Because the COBRE structural images were of higher quality, OK images were selected from COBRE while Maybe and Fail were selected from ADHD-200. For each training image, the rater was first asked to assess the image, and was then able to see the tags and the final ratings by an expert rater (YB).

After completing the training session, raters were presented a series of 100 "open label" cases, and were free to rate as many of these images as they wanted. We chose to present only 100 images in order to ensure we would have many ratings by different raters for each image, within a relatively short time frame. We arbitrarily selected a subset of 100 images with a ratio of 35 Fail, 35 Maybe, and 30 OK images based on one expert rater (YB). Once again, the OK images were drawn from COBRE, while the Fail and Maybe were drawn from ADHD-200.

2.7. Raters

More than 2500 volunteers took part in our Brain Match project. They performed approximately 21,600 ratings of individual images over 2 beta-testing phases and two full workflows for a total of 260 registered brain images (see Brain images section). We used a retirement of 40 ratings, which means each image was rated by 40 different Zooniverse raters before being removed from the workflow. Only individuals who rated more than 15 images were kept in the final study. After data cleaning, 41 Zooniverse volunteer raters were kept. The distribution of rating per image showed a mean number of ratings of 21.76±2.75

(see supplementary material Figure S1).

A group of 9 expert raters were also recruited for this study and each asked to rate all of the 100 validation images using the Brain Match interface. They were instructed to first start with the training session and to carefully read the tutorial before starting the main QC workflow. All raters had prior experience with QC of brain registration in the past. Each rater was free to perform the QC task at her pace without any specific direction on how to do it. The process was completed once all ratings were submitted.

Finally, a radiologist was also recruited for the study. He rated the same 100 images using Brain Match interface, also undergoing the training session before the rating process. Although the radiologist had no prior experience in QC of brain registration, that participant had very extensive experience in examining brain images following a standardized protocol, and served as a gold standard about what to expect from a fully compliant rater, trained on QC solely through available online documentation.

2.8. Agreement statistics

We used Cohen's kappa (Cohen, 1960) to assess inter-rater reliability across all 9 experts (ratings R1 to R9). The kappa metric measures the agreement between two raters who rate the same amount of items into N mutually exclusive categories. The kappa is based on the difference between the observed agreement (p_o, i.e. the proportion of rated images for which both raters agreed on the category) and the probability of chance or expected agreement (p_e). Kappa is computed as follows:

$$k = \frac{p_o - p_e}{1 - p_e} \tag{2.1}$$

In this work we used a weighted kappa metric, which assigns less weight to agreement as categories are further apart (Cohen, 1968). In our QC cases disagreements between OK and Maybe, and between Maybe and Fail count as partial disagreements; disagreements between OK and Fail, however, count as complete disagreements. We used the R package irr (Gamer, 2012) to estimate the weighted kappa and Landis and Koch's (Landis and Koch, 1977) interpretation of the strength of agreement for kappa ≤ 0 =poor , .01 to .20=slight, .21 to .40=fair, .41 to .60=moderate, .61 to .80=substantial, and .81 to 1=almost perfect.

We also used the Sørensen–Dice coefficient (Dice) to assess the agreement within the rating categories of OK, Maybe and Fail (Sørensen et al., 1948), as follows:

$$DSC = \frac{2\,|X \bigcap Y|}{|X| + |Y|} \tag{2.2}$$

where X is the set of images rated "OK" by one rater and Y is the set of images rated "OK" by a second rater, $\bigcap$ is the intersection between two sets, and $|X|$ is the number of images. In plain English, the Dice between two raters for the OK category is the number of images that both raters rated "OK", divided by the average number of images rated "OK" across the two raters. The same Dice measure was generated as well for "Maybe" and "Fail" images. We interpreted Dice coefficients using the same range of strength of agreement as for the Kappa coefficient (≤ 0 =poor, .01 to .20= slight, .21 to .40=fair, .41 to .60= moderate, .61 to .80= substantial, and .81 to 1= almost perfect).

2.9. Consensus panels

We also evaluated the reliability of QC ratings after pooling several raters into a consensus panel. The panel consensus was generated by counting the number of OK, Maybe and Fail attributed to an image from different raters (number of votes). The category with the highest vote count was selected as the consensus on that specific image for the panel. If there was a tie between 2 or 3 categories, the worst category was selected (Fail $<$ Maybe $<$ OK).

We tested different panel configurations, large and small, for expert and Zooniverse raters separately. Large panels were composed either by all 9 experts (panel Ec) or 41 Zooniverse users (panel Zc). We compared the agreement between Ec and Zc versus each individual expert rater (R1 to R9) as well as the ratings from the radiologist (Ra). For small panel, experts were arbitrarily split into three panels of three raters (panels Ec1, Ec2 and Ec3). The Zooniverse users were also arbitrarily split into two independent consensus panels of roughly equal size (Zc1 and Zc2). We quantified the agreement between small panels, as well as small vs large panels.

3. Results

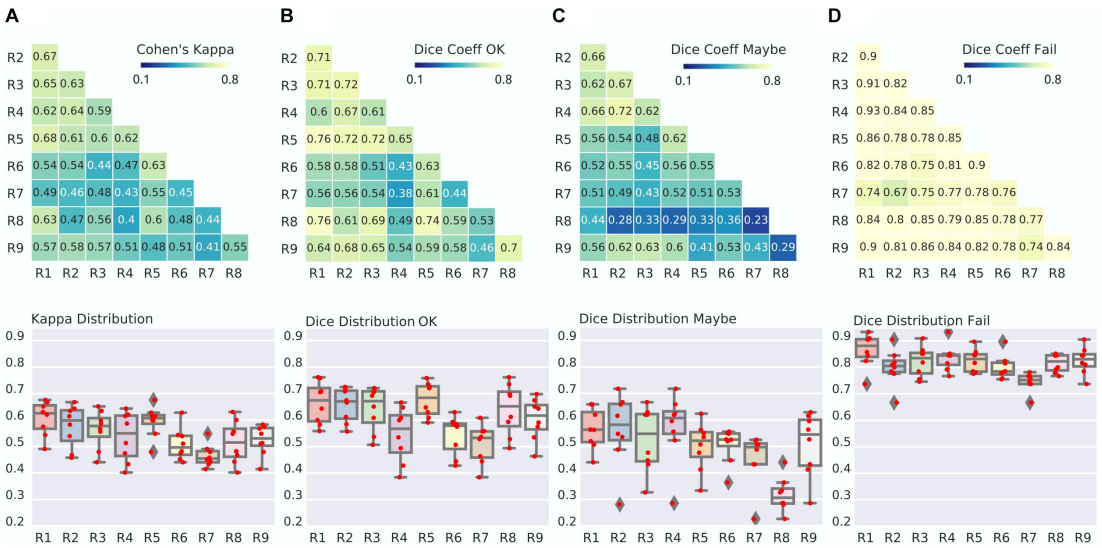


Fig. 7. Between-expert agreement - a: Matrix of Kappa agreement between raters (top). Note that R1 to R9 are identification codes for the different expert raters. The distribution of agreement is also presented (bottom). For example, the boxplot for R1 shows the agreement between R1 and R2-R9. **b-d**: Matrix and distribution for the Dice agreement between raters in the OK (**b**), Maybe (**c**) and Fail (**d**) categories.

3.1. Expert raters achieved moderate agreement, with "Fail" rating being the most reliable

Agreement between expert raters across the three classes (OK, Maybe, Fail) was moderate to substantial (range 0.4 to 0.68, average of 0.54 ± 0.08), see Figure 7. However, there were marked differences in agreement across the three rating classes. The highest reliability was for "Fail", with between-rater Dice agreement ranging from substantial to almost perfect (0.67 to 0.93, average of 0.8 ± 0.06). The second class in terms of reliability was "OK", with Dice ranging from fair to strong (0.38 to 0.76), and the least reliable class was "Maybe", with Dice agreements ranging from slight to strong (0.23 to 0.72).

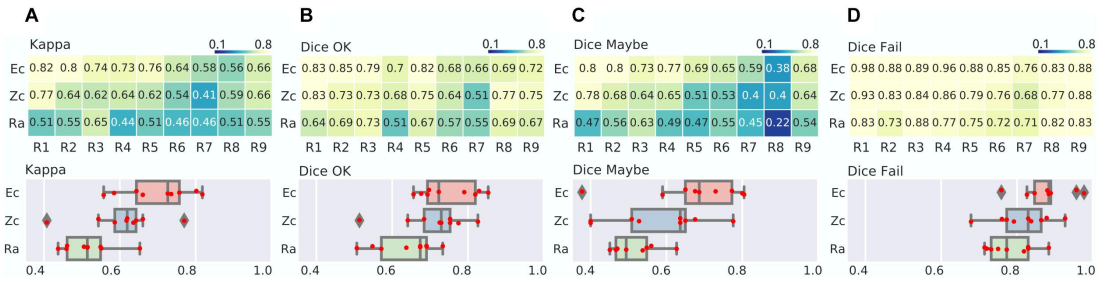


Fig. 8. Zooniverse, expert and radiologist agreements - a: Matrix of Kappa agreement between consensus of experts (Ec), zooniverse users (Zc) and radiologist (Ra) raters, in rows, vs individual experts (R1-R9), in column (top). The distribution of agreement is also presented (bottom). **b-d**: Matrix and distribution for the Dice agreement in the OK (b), Maybe (c) and Fail (d) categories.

3.2. Large panels of expert or zooniverse raters give convergent, reliable QC ratings

We found that the kappa between Ec and individual expert raters was, as expected, improved over comparison between pairs of individual experts, with a range from moderate to strong (0.56 to 0.82), see Figure 8. As observed before, the Dice scores for Ec were highest in the "Fail" category (almost perfect agreement, range of 0.76 to 0.98), followed by the "OK" category (from substantial to almost perfect: range 0.66 to 0.85) and finally "Maybe" (fair to almost perfect, ranging from 0.38 to 0.8). These findings confirmed our previous expert inter-rater analysis, with "Fail" being a reliable rating, "Maybe" being a noisy rating, and "OK" being a moderately reliable rating. When comparing the individual experts with the Zooniverse panel Zc, we only observed a slight decrease in average Kappa compared with the Expert panel (0.61 for Zc vs 0.7 for Ec), mostly driven by the "Fail" (0.82 for Zc vs 0.88 for Ec) and "Maybe" (0.58 for Zc vs 0.68 for Ec) ratings. When directly computing the agreement between the two consensus panels Ec and Zc, the kappa was substantial (0.76), with almost perfect agreement for "Fail" (Dice 0.9) and "OK" (0.82), and substantial agreement for "Maybe" (0.77), see Fig. 8. This comparison demonstrated that aggregating multiple ratings improved the overall quality, and that expert and zooniverse raters converged to similar ratings. The radiologist achieved a level of agreement with panels similar to what was observed with expert raters, and was substantially lower than the agreement between panels. This shows that the QC training material alone was enough for a radiologist to agree with QC experts, but a single user can likely not achieve high quality QC ratings by herself.

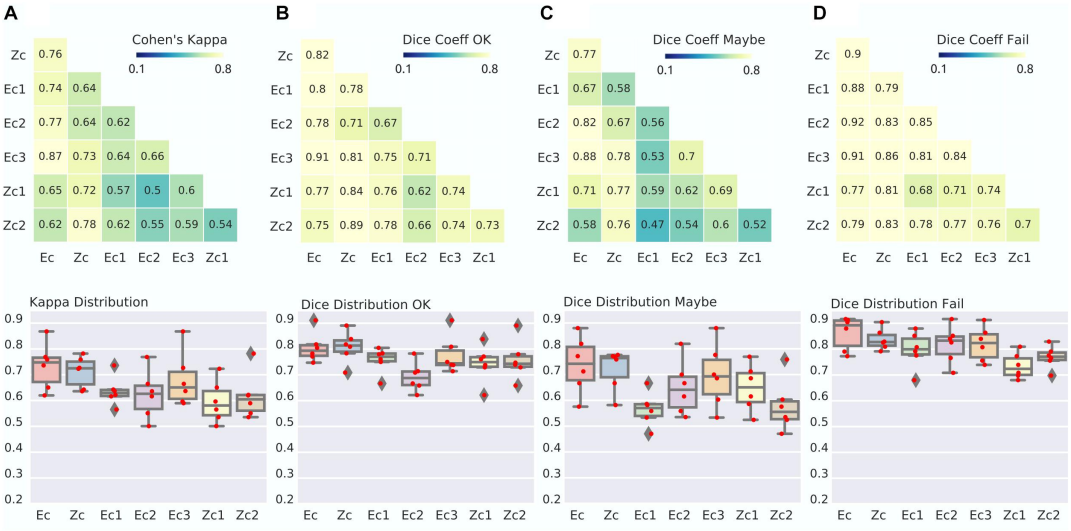


Fig. 9. Agreement between small panels of raters for both experts and Zooniverse panels - a: Matrix of Kappa agreement between large panel consensus of experts (Ec), zooniverse users (Zc) and a small panel of expert (Ec1 = 3 rater, Ec2 = 3 rater, Ec3 = 3 rater) and small panel of Zooniverse raters (Zc1 = 20 rater, Zc2 = 21 rater) (top). The distribution of agreement is also presented (bottom). **b-d**: Dice distribution between group consensus in the OK (**b**), Maybe (**c**) and Fail (**d**) categories.

3.3. Small consensus panels of expert (N=3) or Zooniverse (N=20) raters achieve reliable QC ratings

Once we established that large panels of raters lead to high levels of agreements, our next question was to determine whether small panels could also lead to reliable assessments. The small expert panels Ec1-3 reached lower agreement with Zc than the full Ec. Specifically, kappa was 0.64, 0.64 and 0.73 for Ec1 to Ec3 (with respect to Zc), compared to kappa of 0.76 for Ec vs Zc. Similar observations were made when breaking down the comparison per category with Dice, with a decrease of 5% to 10% in this coefficient (see Figure 9). Comparing small zooniverse panels Zc1-2 with the full expert panel Ec, a slight decrease in reliability was observed, very similar in magnitude with comparisons between Ec1-3 and Zc. The agreements Ec1-3 vs Zc, as well as Zc1-2 vs Ec, remained substantial. This suggests that reliable three-level QC assessments can be performed by small panels of three experts (n=3), or moderate panels of zooniverse users, with roughly 20 assessments by image (see Supplementary Material Figure S01 for distribution).

4. Discussion

This project proposes a standardized QC protocol with minimal training overhead and no required knowledge of brain anatomy. Our goal was to quantify the reliability of QC ratings

between expert raters, as well as panels of expert or Zooniverse raters. Overall, our results demonstrated that our protocol leads to good reliability across individual expert raters, in particular for "Fail" images, and good reliability across panels of raters (both experts and Zooniverse), even for panels featuring only three experts. To our knowledge, this is the first quantitative assessment of between-rater agreement on QC of brain registration.

4.1. Visual QC

Our protocol was designed to be simple enough that even a rater without brain anatomy knowledge or prior QC experience could generate meaningful ratings. The mosaic view of 9 slices used in our protocol is similar to display images used in fMRI preprocessing tools like MRIQC (Esteban et al., 2017), fMRIPrep (Esteban et al., 2018) or CONN (Whitfield-Gabrieli and Nieto-Castanon, 2012). These QC tools also use an overlay that highlights brain borders or tissue segmentation. Differentiating aspects of our protocol are (1) fewer number of brain slices in the mosaic view, so that raters can more easily examine all presented images; (2) the overlay provides an objective confidence interval to assess the severity of misregistration in key anatomical landmarks. We believe that these two design principles helped reduce the subjectivity of brain registration QC, and increase between-rater agreement, although we did not formally test these hypotheses.

4.2. Inter-rater Agreement

Table 1 shows that the visual QC agreement reported in recent studies ranged from 0.39 to 0.9. Interestingly, the studies which reached high levels of agreement (0.93-0.96) used ratings with only two levels (ex: pass, fail). Studies with 3 or more rating levels reported lower agreement scores (0.39-0.85), which were in line with our findings (average of 0.54 for experts). The most challenging rating in our protocol appeared to be the "Maybe" class, featuring mild, spatially limited registration errors. In contrast, good and failed registrations were easily detectable by expert raters. When working with three levels of ratings, the reliability of our protocol is not high enough to work with a single rater. We found that a consensus panel of 3 experts was sufficient to reach a good level of agreement (average of 0.64), which appears as a minimum panel size to generate high quality QC scores. Aggregating rating between expert or non-expert is a good solution to overcome the variability among human observers on the QC task.

4.3. Crowd sourced QC

Crowdsourcing QC rating could be one solution to generate high quality QC ratings in big datasets like the UK biobank (Alfaro-Almagro et al., 2018). A recent work from Keshavan and colleagues (Keshavan et al., 2018b) showed that crowdsourced QC ratings on raw brain

images can reach the performance of an automated state-of-the-art machine learning QC tool (Esteban et al., 2017). This work relied on a large pool (N=261) of participants, many of whom had prior experience in neuroimaging. We recruited more than 2000 zooniverse non-expert raters, and found that a consensus panel of non-experts with adequate size (about 40 ratings per image) leads to QC ratings of similar quality to a panel of three experts.

4.4. Limitations of the study

Our study has a number of limitations. First, our protocol is intended to be used with anatomical brain registration in the context of fMRI analyses in volumetric space, rather than surface. Structural brain imaging studies (i.e. cortical thickness) or surface-based fMRI analyses need other protocols that examine more closely fine anatomy and tissue segmentation. Also, our primary use case is large-scale research studies, and not clinical applications. Some clinical applications may require more stringent standards being applied on brain registration. Our protocol was validated with a specific brain registration tool, the CIVET pipeline, and may not be well suited for other algorithms.

Second, we did not control for screen size, screen resolution or fidelity of colour representations in our validation, be it with experts or zooniverse individuals. The main use case for our protocol is the review of thousands of brain registration (e.g. in the ABCD sample (Casey et al., 2018)) in a relatively short span of time. The quality control procedure only examines coarse anatomical landmarks, and the required precision of the alignment is on the order of a couple of millimetres. For that reason, we think that the characteristics of the screen will not affect significantly between-rater agreement. This is however a potential source of variations which may have decreased the observed agreement, both between experts and zooniverse raters.

Third, the success rate of our registration tool varies widely as a function of the imaging protocol. The Cobre dataset has almost only OK registration, while the ADHD has a lot of Maybe and some Fail. So we decided to mix two datasets, in order to assemble representative examples of the three classes. This may influence the results by increasing the potential agreement if subjects learned to recognize which datasets the examples originated from.

Fourth, our choice on the number of rated images (N=100) was selected arbitrarily. We checked the appropriateness of that choice by assessing the minimum number of rated cases with a 3-choice decision using the R package "irr" (Gamer, 2012), that uses the minimum sample size estimation formula from Flack and colleagues (Flack et al., 1988). We estimated

the minimum sample size under the following scenario. The vector of marginal probability was given by rates for the 3 categories, OK = 0.3, Maybe =0.35 and Fail =0.35. These marginal probabilities were decided by our team when designing the dataset, based on an initial QC assessment performed by YB and PB. The value of kappa under the null hypothesis was set equal to 0.5 (k0=0.5) - i.e. we want to demonstrate an improvement over a baseline kappa of 0.5. The true kappa statistic estimated between two experts was set equal to 0.72 (k1=0.72), as was observed in our sample. The type I error test was set equal to 0.05 (alpha=0.05). The desired power to detect the difference between the true kappa and the null kappa was investigated at 0.8 and 0.9, separately. The required number of ratings was estimated at N=54 for a power of 0.8, and N=72 for a power of 0.9. In our case, the number of images rated per expert was N=100, which is more than required by the power analysis.

Fifth, we were unable to access to what degree this protocol improves or not over current best practices in the fMRI community, in the absence of other standardized protocols available for comparison. We still produced preliminary evidence while developing the current protocol. During the beta phase of our project, we tested the agreement between consensus of Zooniverse raters and expert raters (on 29 images). The protocol used during that phase was different from the present one. In particular, we did not instruct raters on how to take the final decision on the quality of registration (Figure 6 c-e), and we did not offer a training set. The kappa measure between consensus Zooniverse raters and an expert during phase 1 was 0.34, by contrast with 0.61 using the current protocol. We regard these results as preliminary evidence that our protocol improves over our previous iteration. These results are to be interpreted with caution, as the number of images rated was low and we used only one expert rating. Note that the feedback received by beta testers helped us identify the importance of steps described in Figure 6 c-e, and we suspect that protocols that do not include such detailed explanation have poor reliability. But we did not attempt to demonstrate this formally within the scope of the present study.

Finally, our protocol is missing an evaluation of another key registration step, i.e. alignment between functional images and the structural scan (Calhoun et al., 2017). We are currently working on an extension of our protocol for functional registration.

4.5. Future work: impact of QC on downstream analyses

Despite the ubiquity of visual brain registration QC in the neuroimaging research community, the impact of visual QC of brain registration on statistical analyses remains poorly characterized. Gilmore and colleagues (Gilmore et al., 2019) used a multi-site dataset of

structural MRI images with different age ranges to show how automated image quality metrics impacted regional gray matter volumes and their relationship with age. Ducharme and colleagues (Ducharme et al., 2016) showed a significant impact of visual QC on the estimation of cortical trajectories. They demonstrated that, when omitting to discard subjects that did not pass QC, the developmental trajectory of cortical thickness followed a quadratic or cubic trend. By contrast, after filtering those subjects, the trajectory followed a linear trend. Standardizing the QC protocol will allow different laboratories to join their effort of rating and open up new opportunities to systematically investigate the impact of visual QC on the relationship between the brain and various phenotypes. This represents an important area of future work for brain registration.

5. Conclusion

Our QC protocol is the first reliable visual protocol for brain registration in fMRI studies. The protocol is easy to implement and requires minimum training effort. This protocol demonstrates a good reliability when performing a two-level assessment (Fail vs OK/Maybe) by an individual rater, or aggregating multiple three-level ratings (OK, Maybe, Fail) from a panel of experts (3 minimum) or non-experts (15 minimum). The images necessary to apply the protocol can be generated using an open-source tool, called dashQC,_fmri (Urchs et al., 2018) and a live version can be tested on this link https://simexp.github.io/dashQC_BrainMatch/index.html. We hope this new protocol will help standardize the evaluation and reporting of brain registration in the fMRI community. This standardization effort will also enable the generation of high quality QC ratings on large amounts of data, which will in turn allow training machine learning models to automatically perform brain registration QC, alleviating the need for visual reviews.

6. Contribution to the Field Statement

Automatic alignment of brain anatomy in a standard space is a key step when processing magnetic resonance imaging for group analyses. Such brain registration is prone to failure, and the results are therefore typically reviewed visually to ensure quality. There is, however, no standard, validated protocol available to perform this visual quality control. In this work, we proposed the first standardized and validated QC protocol for brain registration. We hope this new protocol will help standardize the evaluation and reporting of brain registration in the fMRI community. This standardization effort will also enable the generation of high quality QC ratings on large amounts of data, which will in turn allow training machine learning models to automatically perform brain registration QC, alleviating the need for visual reviews.

7. Acknowledgement

We would like to acknowledge the following people for their contributions to the project. The 9 expert raters: Perrine Ferre, Aman Badhwar, Angela Tam, Jeni Chen, Cole Zweber, Sebastian Urchs, Carolina Makowski, Elizabeth DuPre and Paule Joanne Toussaint. Also we thank Andrew Doyle, Sara Dupont, Benjamin De Leener and Maxime Pelland for their contribution on the beta testing phase. We would like to thank the Zooniverse team, especially Helen Spiers and Laura Trouille, for their feedback and support along the process. Finally, we thank all the citizen-scientists on Zooniverse who participated in our project Brain Match. We are very grateful for your contributions.

Supplementary material

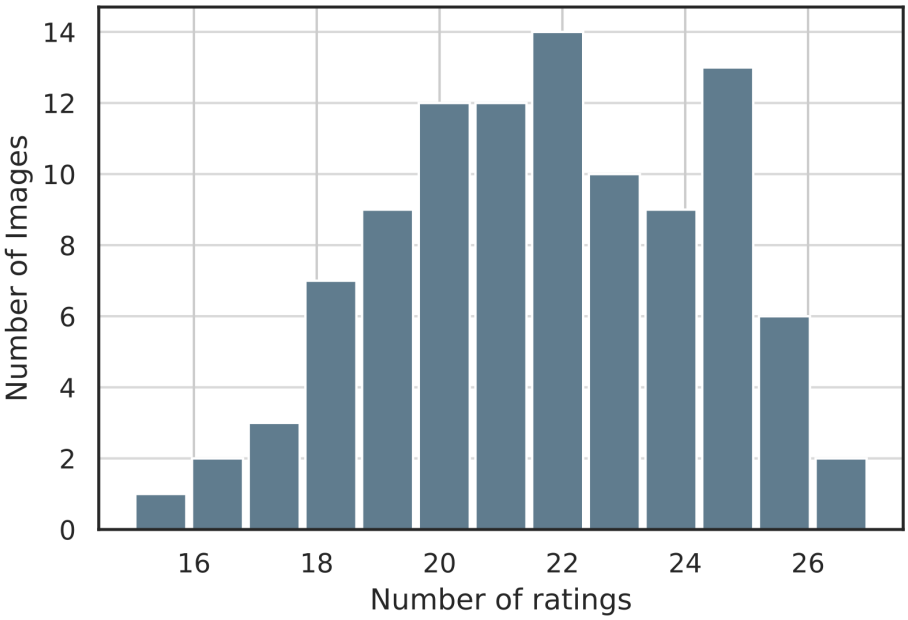


Fig. S 10. Distribution of images as a function of the number of ratings (mean of 21.76 ± 2.75, range 15 to 27).

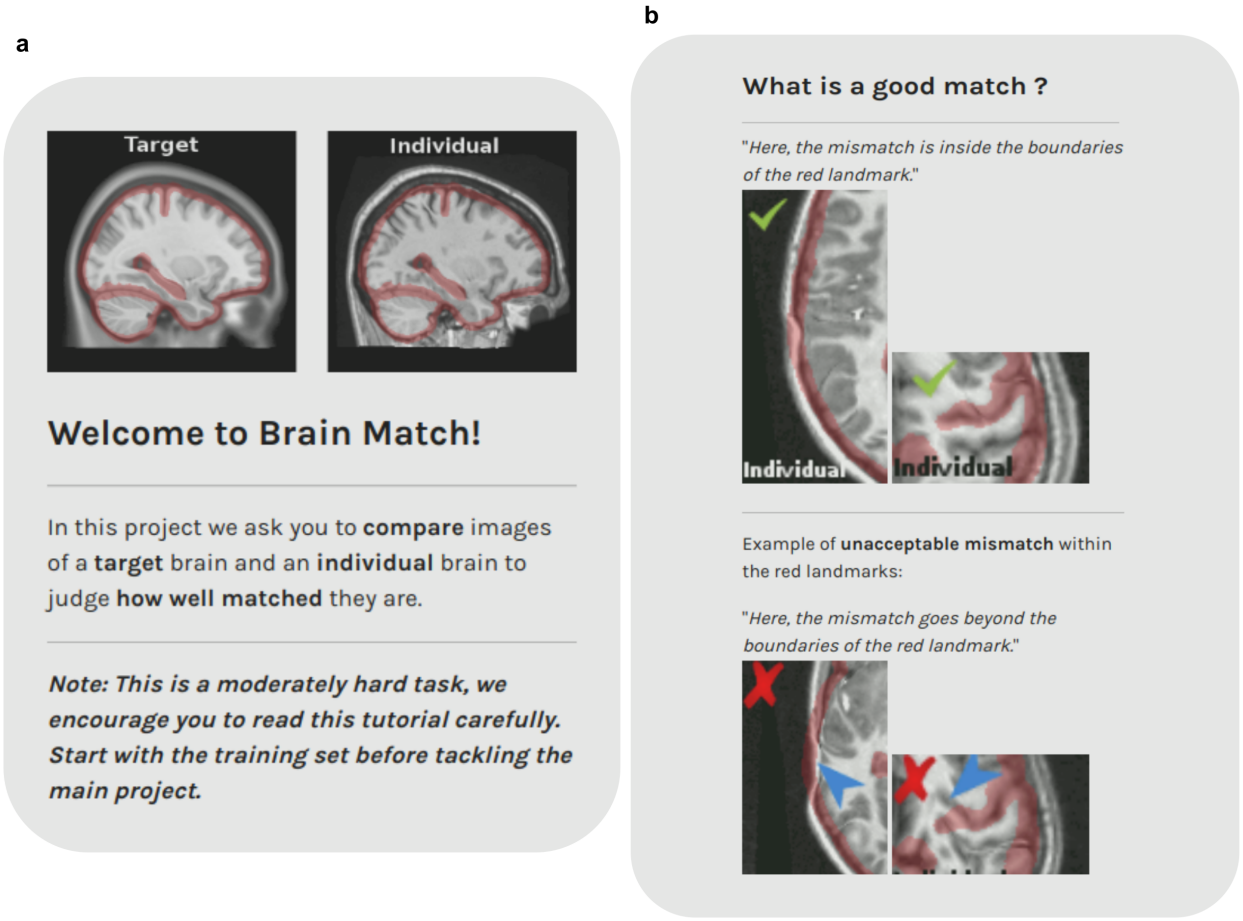


Fig. S 11. Brain Match tutorial interface - **a**: A tutorial popup when first accessing the brain match interface showing the main goal of the task. **b**: A panel of annotated examples of what is a good and bad match between highlighted brain areas. More examples follows on how to tag images and how to decide the final rating of each image (for more detailed tutorial visit brain-match)

Second Article.

Subtypes of brain activation are heritable and genetically linked with behaviour in the Human Connectome Project sample

by

Yassine Benhajali[1], AmanPreet Badhwar[2], Urchs Sebastian[3], Moreau Clara[4], Chouinard-Decorte Francois[5], Vainik Uku[6], Ferre Perrine[7], Orban Pierre[8], Daniel Pérusse[9], and Pierre Bellec[10]

([1]) Département d'anthropologie, Université de Montréal, Montréal, Québec, Canada.

([2]) Centre de recherche de l'Institut universitaire de gériatrie de Montréal (CRIUGM), Montréal, Québec, Canada.

([3]) Montreal Neurological Institute, Montréal, Québec, Canada

([4]) CHU Sainte-Justine Research Center, Montréal, Québec, Canada.

([5]) Montreal Neurological Institute, Montréal, Québec, Canada.

([6]) Institute of Psychology, University of Tartu, Estonia

([7]) Centre de recherche de l'Institut universitaire de gériatrie de Montréal (CRIUGM), Montréal, Québec, Canada.

([8]) Centre de recherche de l'Institut universitaire en santé mentale de Montréal.

([9]) Département d'anthropologie, Université de Montréal, Montréal, Québec, Canada.

([10]) Département de psychologie, Université de Montréal, Montréal, Québec, Canada.

This article was submitted in Neuroimage.

Abstract.

Many imaging and genetic studies have aimed to clarify whether the brain acts as an intermediate phenotype mediating the influence of genes in human behaviour. Brain activation in response to task demand are heterogeneous at the individual level, but also follow common patterns at the group level. Some studies have addressed this tension between heterogeneity and homogeneity by identifying groups of individuals that share the same brain activation patterns, called brain activation subtypes.

In this work, we aimed to assess the viability of brain subtypes as endophenotypes between genes and behaviour. We extracted brain activation subtypes separately for seven fMRI tasks, in 842 participants from the Human Connectome Project (HCP). We estimated the heritability of these subtypes and their genetic correlation with behavioural measures obtained inside and outside the scanner.

Across all tasks, subtypes ranged from a predominantly 'deactivating' pattern toward a more 'activating' pattern of brain activity, with heritability estimates ranging from 0 to 0.62. We observed high genetic and phenotypic correlation between behavioural measures and brain activation subtypes only for language and working memory tasks.

Our results showed a significant genetic grounding of brain activation subtypes and they appear as a simple yet effective technique to tackle heterogeneity into imaging genetic studies.

Keywords: Task-fMRI, Subtypes, Heritability, Genetic Correlation, Behavioural measures.

1. Introduction

Human actions and interactions are influenced by a hierarchy of factors such as genetic make-up, environmental influences, culture, individual values and attitudes. Within this hierarchical system, functional brain organization is an intermediate phenotype, and could be the key to understanding how factors of nature and nurture drive variations in human behaviour. Activation maps generated using non-invasive functional magnetic resonance imaging (fMRI) offer a window into the individual functional organization for a specific cognitive context. It is well established that these activation maps exhibit substantial variability across individuals (Van Horn et al., 2008), which may explain inter-individual differences in observable behaviour. However, it is still unclear how this functional brain variability relates to behavioural differences in terms of genetic and environmental influences. The main

objectives of the present study were to (1) estimate the heritability of brain activation subtypes, i.e. groups of individuals who reliably share similar activation maps and, (2) to assess whether these subtypes share their genetic basis with behavioural measures across a broad range of cognitive tasks.

1.1. Inter-individual differences in brain activation

Task fMRI activation maps exhibit substantial inter-individual differences that likely reflect both fundamental differences in the anatomo-functional organization of brain networks, and differences in cognitive strategies. An early electrophysiological study comparing proficient and poor readers showed that proficient readers had more efficient neural processes, as reflected by lower power spectra in EEG (Maxwell et al., 1974). In a more recent fMRI study using a sentence-picture verification paradigm, Miller and colleagues (Miller et al., 2012) manipulated the use of either verbal or visual strategies in a memory retrieval task in healthy subjects. They found that larger differences between two individuals in their tendency to visualize highly imageable word stimuli were associated with larger differences between their patterns of brain activity across the whole brain. Differences in brain activation related to either cognitive strategies or individual abilities have been demonstrated across a wide range of cognitive domains (Heun et al., 2000; Machielsen et al., 2000; McGonigle et al., 2000; Miller et al., 2009, 2002; Feredoes and Postle, 2007; Seghier and Price, 2009; Seghier et al., 2008; Parasuraman and Jiang, 2012). Such inter-individual differences in brain activation have not only been observed for complex cognitive processes but even for simple tasks. For example, Seghier and colleagues 2016 found that the primary motor cortex was missing from the group activation map in a motor task, but was visible in nearly 50% of individual activation maps. Taken together, these observations show that average group activation maps do not accurately represent the neural systems engaged by the individuals that make up that group (Lebreton and Palminteri, 2016).

1.2. Brain subtypes

Clustering strategies such as hierarchical clustering have the potential to study the inter-individual heterogeneity in brain activation. A cluster analysis summarizes the heterogeneity of brain maps using a finite number of subtypes (See Figure 1). Hierarchical clustering was for instance applied to aggregate subjects based on the similarity of their brain activation maps (Kherif et al., 2003), and to identify atypical subjects (Seghier et al., 2007). Clustering-based subtyping has recently gained popularity for the dissection of brain heterogeneity, for example the delineation of cortical atrophy subtypes in Alzheimer's disease (Badhwar et al., 2019a), and resting-state connectivity subtypes in the autism spectrum (Easson et al., 2019; Urchs et al., 2020). In the non-clinical domain, some studies

have linked brain functional subtypes with normal behavioural and demographic measures. A recent study by Kashyap and colleagues (Kashyap et al., 2019) investigated resting-state connectivity subtypes of 788 healthy subjects from Human Connectome Project (HCP). They estimated the common component of each subject from the four resting states runs, and classified the subjects based on the similarity/dissimilarity of their components. Two subgroups emerged, one with high similarity and another with high dissimilarity between components. Both subgroups associated with behavioural measures related to the use of marijuana, illicit drugs, alcohol, and tobacco, and antisocial personality. Another study by Smith and colleagues (Smith et al., 2015) showed that a positive-negative mode of population covariation links brain connectivity, demographics, and behaviour. Finally, a study by Kirchhoff and Buckner (Kirchhoff and Buckner, 2006) reported that different cognitive strategies (visual vs verbal) used by subjects in the scanner to perform a memory task were correlated with distinct brain activation networks. They also found that activity in brain regions associated with the use of these strategies was also correlated with memory performance. These studies demonstrate that heterogeneity in brain function exists and may be characterized as subtypes using unsupervised data mining tools.

Despite the growing literature on resting-state-based brain subtyping, only few studies explored the potential use of task-based brain subtyping (activation subtpes). Therefore, it is unclear if activation subtypes could serve as a relevant intermediate phenotype between genes and behaviour. In particular, are brain activation subtypes genetically driven? And, if yes, are the same genes driving inter-individual variations in behaviour and brain subtypes?

1.3. Genetic of brain activation subtypes

A quantitative phenotype, or trait, depends on the cumulative actions of many genes and of the environment. The quantitative genetic analysis uses either family-based quantitative data or twins-based data to partition the observed covariance among related individuals into genetic and environmental components. Heritability is defined as the contribution of additive genetic variance to the total variance of a trait (Falconer et al., 1996) and can be estimated for any quantitative phenotype. A meta-analysis from imaging genetics showed that variations in neuroimaging phenotypes, and in particular activation maps, are substantially (ex: up to 0.75% on visuomotor task) explained by additive genetic factors (Blokland et al., 2012). Current methods in quantitative genetics to estimate the heritability of fMRI activation maps, however, suffer from important limitations. The large number of voxel- and vertex-wise measurements in imaging genetics make heritability estimates very challenging in terms of computational load and the need to account for multiple testing (Ganjgahi et al., 2015). Studies that estimated heritability of brain activation tended to

mitigate multiple comparisons by targeting regions of interest (Blokland et al., 2008, 2011, 2017; Rao et al., 2018; van der Meulen et al., 2018; Chen et al., 2018; Babajani-Feremi, 2017; Pinel and Dehaene, 2013). To our knowledge, there is no study that investigated heritability of brain activation subtypes in healthy samples. Investigations of the heritability of brain activation maps will thus benefit from the subtyping approaches reviewed earlier, as they provide compact summaries of the variations in brain activity across subjects, making it easier to interpret what specific characteristics of activation maps are under genetic control.

Some studies attempted to estimate whether brain activation maps and behavioural measures are influenced by similar groups of genes. This concept is called pleiotropy or genetic correlation between phenotypes. Le Guen and colleagues (Le Guen et al., 2018) used HCP data to perform bivariate genetic analyses between the fMRI activation maps and fluid intelligence, as well as the performance of subjects on working memory and language domain tasks. Regions of the language network along the superior temporal sulcus and the angular gyrus were significantly genetically correlated with indicators of cognitive performance (fluid intelligence, working memory, vocabulary comprehension and reading decoding). Rao and colleagues (Rao et al., 2018) also provided evidence for the genetic correlation between risk-taking behaviour and risk-related brain activation. Those studies used a voxel-level analysis on regions of interest in order to estimate the genetic correlation between brain phenotypes and behavioural phenotypes. In this work we used subtypes weights as brain phenotype to capture genetic link with behavioural measures.

1.4. Aims and hypotheses

In this work, we aimed to demonstrate the viability of activation subtypes as an intermediate endophenotype between genes and behaviour. We implemented a clustering approach to identify subtypes of brain activation in the fMRI tasks of the HCP sample. Our two main objectives were 1) to test the heritability of these brain activation subtypes 2) to test the association and shared genetic variance between brain activation subtypes and a battery of behavioural measures. The clustering technique proceeded by aggregating individual brain activation maps into a limited number of subtypes. The similarity between each individual map and each subtype was then estimated with a continuous subtype weight. Using these subtype weights as a trait, we specifically aimed to:

(1) Test the association of subtypes weights with behavioural measures, collected outside of the scanner, as well as in-scanner task performance. Our hypothesis was that the performance of each task would be associated mainly with the brain activation

subtypes derived from this task, while behavioural measures would be associated with tasks drawn from a related cognitive domain.

(2) Evaluate if these activation subtypes are heritable, using a twin family design. Our hypothesis was that subtype of brain activation would show moderate to high heritability.

(3) Quantify the genetic correlation between subtypes of brain activation and behavioural measures in several domains, as well as with the in-scanner task performance. Our hypothesis was that these genetic correlations would be moderate to high between behavioural and brain phenotypes that were related to similar cognitive abilities.

2. Method

2.1. Subjects

We used the S900 data release from the HCP open access database (Van Essen et al., 2013). The S900 release includes behavioural and imaging data from 897 healthy adult participants collected August 2012–March 2015, out of which 862 healthy adults (aged 22–35 y) successfully completed at least one of the 7 fMRI tasks. Post quality control, 842 subjects were selected for the analysis - included 278 families (742 individuals), and 100 unrelated subjects. Families were either monozygotic twins (genetically identical, 84 pairs, 168 individuals) or dizygotic/fraternal twins (genetically no more related than ordinary full siblings, 77 pairs, 154 individuals) and 89 families of non-twin siblings (209 individuals) (Van Essen et al., 2013).

The HCP sample reflects the ethnic diversity of American families (White non-Hispanic, Hispanic, Asian and African-American). HCP provides the required ethics and consent needed for study and dissemination. This study went through further internal institutional data analysis approval.

2.2. Imaging data

MRI data for the S900 release was collected on a customized Siemens MAGNETOM Connectom 3T scanner at Washington University. The data is composed of multiple imaging sessions covering four modalities: structural (T1w and T2w), resting-state fMRI , task fMRI (tfMRI, 7 tasks) and diffusion MRI (dMRI) (Hodge et al., 2016). In this work we only used T1w and tfMRI modalities. For T1w images a 32 min was spent acquiring each MPRAGE image. An MPRAGE sequence with 0.7 mm isotropic resolution (FOV = 224 mm, matrix = 320, 256 sagittal slices in a single slab), TR = 2400 ms, TE = 2.14 ms, TI = 1000 ms, FA = 8°, Bandwidth (BW) = 210 Hz per pixel, Echo Spacing (ES) = 7.6 ms, see (Glasser

et al., 2013) for more detail. For each subject and task, 2 runs of tfMRI time series data were collected, 1 run in each phase encoding direction, left-to-right (LR) then right-to-left (RL), with a temporal resolution of 0.73 s and a spatial resolution of 2-mm isotropic, using a standard 32-channel Siemens coil (Van Essen et al., 2013).

All participants included completed 7 tasks in the MRI scanner - Working Memory (405 frames/scan), Gambling (253 frames/scan), Motor (284 frames/scan), Language (316 frames/scan), Social Cognition (274 frames/scan), Relational Processing (232 frames/scan), and Emotion Processing (176 frames/scan) (Hodge et al., 2016).

2.3. Image preprocessing

From HCP S900 data release, we used the unprocessed tfMRI data. Each subject was preprocessed using the NeuroImaging Analysis Kit version dev-0.14.0 (NIAK[3]), under CentOS with Octave[4] version 3.8.1 and the Minc toolkit[5] version 0.3.18. Analyses were executed in a high-performance computing environment[6], using the pipeline system for Octave and Matlab (PSOM) (Bellec et al., 2012).

Each tfMRI image was corrected for differences in timing of slice acquisitions; a rigid-body motion was then estimated using Minctracc (Collins and Evans, 1997) for each time frame, both within and between runs, as well as between one fMRI run and the T1 scan for each subject. The T1 scan was itself non-linearly co-registered to the Montreal Neurological Institute (MNI) ICBM152 stereotaxic symmetric template (Fonov et al., 2011a), using the CIVET pipeline (Ad-Dab'bagh et al., 2006). The rigid body, fMRI-to-T1 and T1-to-stereotaxic transformations were all combined to resample the fMRI in MNI space at a 3 mm isotropic resolution. To minimize artifacts due to excessive motion, all time frame showing an average frame displacement (FD) greater than 0.5 mm were removed (Power et al., 2012). The following nuisance covariates were regressed out from the tfMRI time series: slow time drifts (basis of discrete cosines with a 0.01 Hz high-pass cut-off), average signals in conservative masks of the white matter and the lateral ventricles as well as the first principal components (accounting for 95% variance) of the six rigid-body motion parameters and their squares (Giove et al., 2009). Finally, the tfMRI volumes were spatially smoothed with a 6 mm isotropic Gaussian blurring kernel. A more detailed description of the pipeline can be found on the NIAK website[7].

[3]http://niak.simexp-lab.org/
[4]https://www.gnu.org/software/octave/
[5]http://bic-mni.github.io/
[6]https://www.calculquebec.ca/
[7]http://niak.simexp-lab.org/build/html/PREPROCESSING.html

2.4. Registration and motion quality control

We used a structured visual quality control procedure (Benhajali et al., 2020a) to assess the registration between the individual T1 scans and the stereotaxic template space, as well as registration between the individual T1 and the median functional image. The quality control procedure included a series of key anatomical landmarks and associated confidence intervals. The quality control procedure is publicly available as a tutorial on the Zooniverse platform (noa, 2015a). Subjects were excluded from the analysis in cases of failed coregistration with the template space or were fewer than 100 time points had acceptable amounts of head motion after scrubbing (see next section). A total of 843 subjects (477 female) passed quality control criteria for inclusion in the analysis.

In scanners head motion is a critical confounding factor in functional MRI and tends to act as a relatively stable, heritable phenotype (Couvy-Duchesne et al., 2014; Engelhardt et al., 2017; Couvy-Duchesne et al., 2016). Removing time points affected by large head motion, a technique known as scrubbing, is efficient at reducing the effect of head motion on fMRI time series (Caballero-Gaudes and Reynolds, 2017). Average framewise displacement (FD) (Power et al., 2012) is a summary measure of head displacement throughout an fMRI acquisition, and can be estimated both before and after scrubbing. We assessed the heritability of average FD motion measure before and after removal of time frames with excessive motion (time frames with FD greater than 0.5 mm). Heritability analysis was implemented using the SOLAR Eclipse software, as described in the dedicated Methods section below. Figure Sup 18 (supplementary material) shows that FD measures before scrubbing had substantial (and significant) heritability estimates, ranging from 0.22 to 0.52 across tasks, which is consistent with previous reports (Couvy-Duchesne et al., 2014; Engelhardt et al., 2017; Couvy-Duchesne et al., 2016). After scrubbing, the heritability estimates of FD ranged from 0 to 0.13 across tasks. We concluded that using scrubbing as a preprocessing step was beneficial for further heritability analysis of fMRI activation maps, as it reduced the potential confounding effects of heritable motion artifacts.

2.5. Task fMRI

2.5.1. Working Memory. N-back tasks were used as follows: the participant was presented with a series of pictures, drawn from different categories (faces, tools, places and body parts). Half of the blocks were used for a 2-back memory trial, where the participant was asked to respond each time whether the current stimulus was the same as the one from two back. The other half of the blocks were used for a 0-back working memory task, where the individual had to recognize a single 'target' for any stimulus during the block.

2.5.2. Incentive Processing (gambling). Participants played a game of card guessing, and were asked to guess the number on a mystery card (question mark card) to win or lose money. The potential number on the mystery card was an integer between 1 and 9. Pressing one of two buttons, the participants had to guess if the number was over or below 5. A successful guess was rewarded by 1$, the participants lost 0.50$ for bad guesses, and nothing for neutral guesses (where the number was exactly 5). The task was presented in blocks of 8 trials that were either mostly reward (6 reward trials pseudo randomly interleaved with either 1 neutral and 1 loss trial, 2 neutral trials, or 2 loss trials) or mostly loss (6 loss trials interleaved with either 1 neutral and 1 reward trial, 2 neutral trials, or 2 reward trials). In each of the two runs, there were 2 mostly reward and 2 mostly loss blocks, interleaved with 4 fixation blocks (15 s each)

2.5.3. Motor. Participants were presented with visual cues that asked them to tap their left or right fingers, squeeze their left or right toes, or move their tongue, in order to generate maps of different motor areas. Each block of a movement type lasted 12 s (10 movements), and was preceded by a 3 seconds cue. In each of the two runs, there were 13 blocks, with 2 blocks of each body part and three blocks of 15 s fixation.

2.5.4. Language Processing. The task consisted of two runs that each interleave 4 blocks of a story condition and 4 blocks of a math condition. The task was designed so that the math blocks match the length of the story blocks, without fixation between block transitions. The story blocks presented participants with brief auditory stories (5–9 sentences) adapted from Aesop's fables, followed by asking participants to select the topic of the story among two alternatives (button press response). The math blocks required subjects to complete addition and subtraction problems (auditory instructions, and select the right answer among two alternatives (button press response).

2.5.5. Social Cognition (Theory of Mind). Participants were presented with short video clips (20 s) of objects (squares, circles, triangles) either interacting in some way, or moving randomly. Participants were asked to judge whether a social or a random interaction was happening between these objects on the video. Each of the random and social conditions had 5 video blocks (2 Mental and 3 Random in one run, 3 Mental and 2 Random in the other run) and 5 fixation blocks (15 s each).

2.5.6. Relational Processing. In the relational processing condition, participants were presented with 2 pairs of objects, with one pair at the top of the screen and the other pair at the bottom of the screen. They were told that they should first decide what dimension differed across the top pair of objects (shape or texture) and then they should decide whether the bottom pair of objects also differed along that same dimension.

2.5.7. Emotion Processing. Participants were presented with blocks of trials that either asked them to decide which of two faces presented on the bottom of the screen matched the face at the top of the screen, or which of two shapes presented at the bottom of the screen matched the shape at the top of the screen. The faces had either angry or fearful expressions.

For more details on tfMRI in the HCP, see Barch and colleagues (Barch et al., 2013).

2.6. Activation maps

For each of the 7 tasks, individual activation maps were extracted from block design conditions using the statistical analysis toolbox FMRISTAT (Worsley et al., 2003). Different conditions were included in some of the tasks (e.g. foot, hand and tongue for the motor task), resulting in a total of 21 conditions. For each of the 7 tasks, the stimuli in the design matrices were convolved with a hemodynamic response function modelled as a difference of two gamma functions. Temporal drift was removed by adding a cubic spline in the frame times to the design matrix, and spatial drift was removed by adding a covariate in the whole volume average. The autocorrelation parameters were estimated at each voxel and used to whiten the data and design matrix.

2.7. Subtypes of activation maps and weights extraction

For each of the 21 task conditions, a nuisance regression was applied on stacked activation maps (Figure 12a). The nuisance aggressors are frame displacement (FD), systolic blood pressure, diastolic blood pressure (Sys Dias) and body mass index (BMI) (Figure 12b). Then, a subject by subject similarity (Pearson's correlation) matrix summarized the between-subject similarity of activation maps (Figure 12c). Next, a hierarchical cluster analysis was applied on all individual activation maps, which identified homogeneous subgroups of maps for each condition. Within each activation map subgroup, we estimated the average activation map across all subgroup members, or subtype map (Figure 12d).

Finally, we computed the spatial similarity (correlation) of each individual activation map in the discovery sample with each of the identified activation subtypes (Figure 12e). The estimated spatial correlation coefficient was called the 'subtype weight' and each individual had a subtype weight for each subtype, ranging from -1 (perfect anticorrelation of the individual and the subtype map) to +1 (perfect correlation of the individual and subtype map). The subtype weights represent a continuous measure of similarity between a given individual activation map and a given subtype map. Unlike discrete assignments made by the hierarchical clustering, the weights are a soft assignment and can be seen as a dimension reduction of the individual maps, rather than pure categorical summary. It has been argued

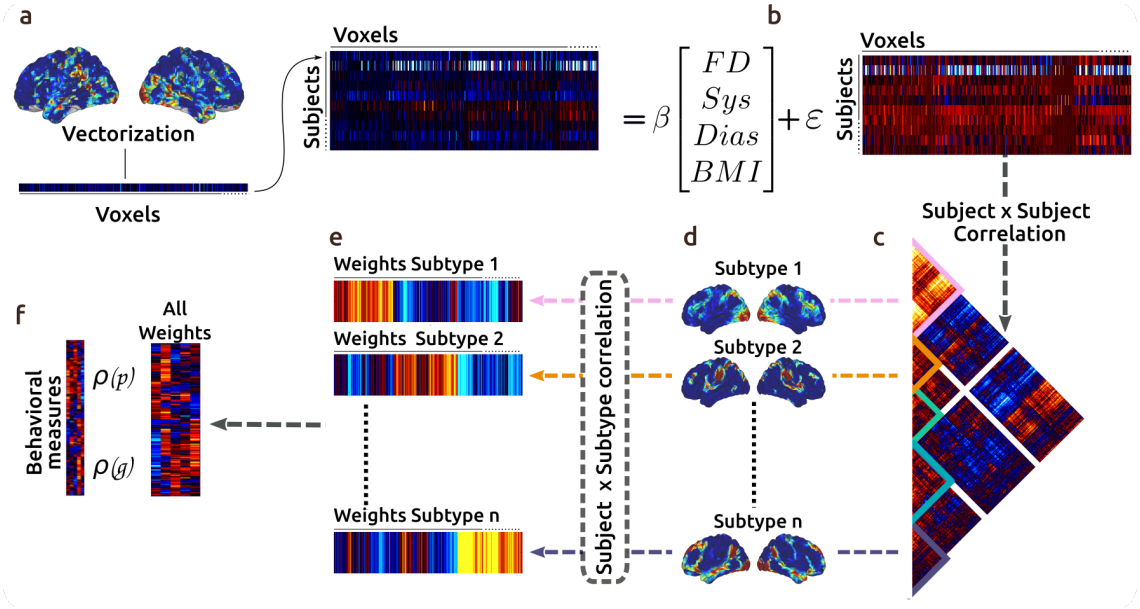


Fig. 12. Activation subtypes maps generation and weights extraction- a. From each subject, an un-thresholded statistical parametric map (SPM) is vectorized then stacked. **b**. Nuisance regression from stacked SPM (FD: Frame Displacement, Sys: Systolic blood pressure, Dias: Diastolic blood pressure, BMI: Body Mass Index). **c**. From the residual stack, a subject by subject correlation matrix is clustered into subgroups of activation maps (subtypes). **d**. Each subtype is represented spatially by a mean activation map of all subjects within the same cluster . **e**. Each subtype is represented on the subject level by its weight (correlation between a vectorized subject activation map and a vectorized subtype map). **f**. Individual weights are correlated genetically (ρ_g) and phenotypically (ρ_p) with behavioural measures.

by several prior studies that such continuous summaries are more appropriate than discrete assignments (Zhang et al., 2016; Badhwar et al., 2019b), and we used the subtype weights as our primary metric for all subsequent statistical analyses (Figure 12f).

2.8. Clustering Behavioural measures

The majority of out-of-scanner HCP behavioural measures were composed of items from the NIH toolbox (noa, 2015b). The behavioural measures were categorized into several domains that included measures of alertness, cognition, emotion, personality, motor and sensory processes in healthy individuals for a total of 74 items (supplementary figure Sup 19 gives details of individual behavioural measures). Since behavioural measures belonged to different categories, we reduced the number of items from 74 into 13 categories using a K-means clustering technique. Because participants were not statistically independent, but clustered into families, we first extracted the correlation between each behavioural measure using the Solar Eclipse polygenic model[8]. This model allowed us to control for the effect of the family,

[8]http://solar-eclipse-genetics.org/

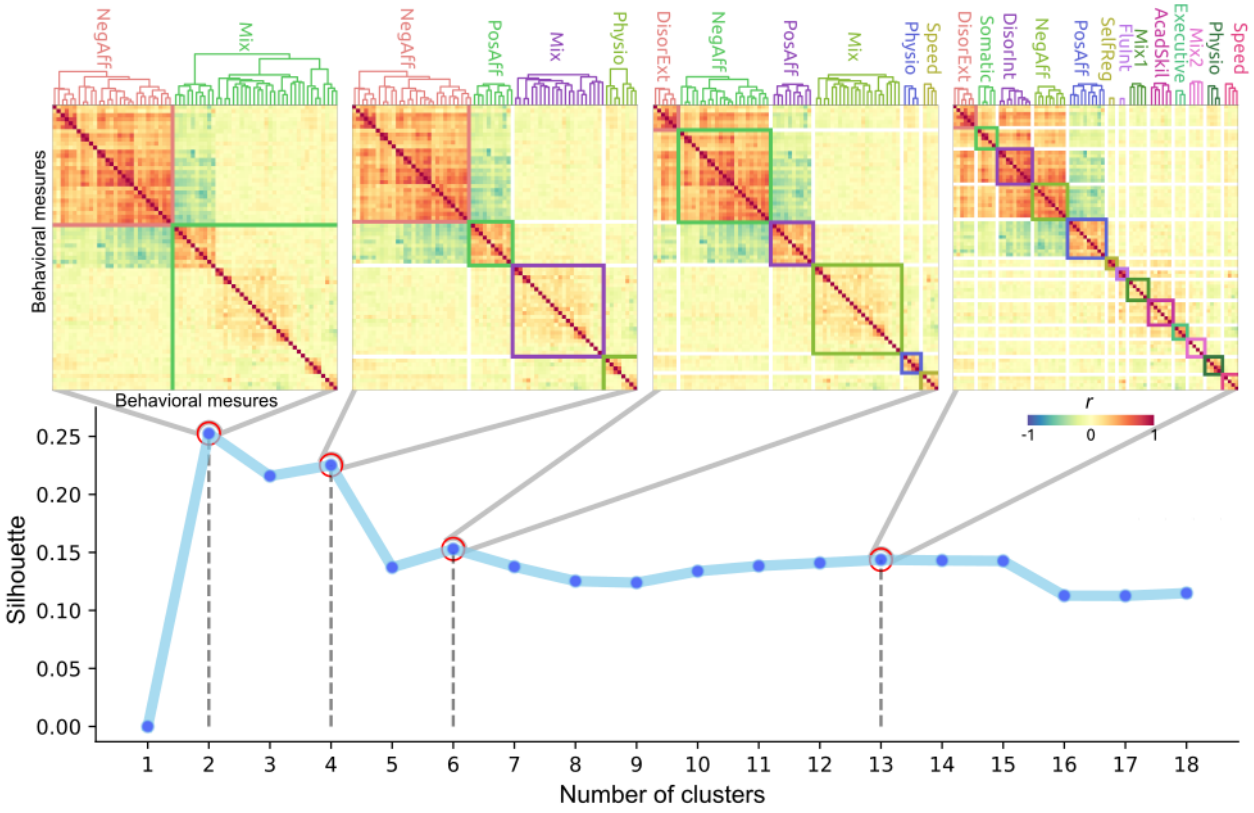


Fig. 13. Behavioural dimensionality reduction- a. A correlation matrix of a collection of 74 behavioural tests are clustered using k-means. The Silhouette plot shows 4 local maxima of cluster separation. The 13 clusters cut off are **Speed**: task reaction time measures, **Physio**: physiological measures, **Mix2** and **Mix1**: a mixed category of measures, **Executive**: working memory and cognitive flexibility measures, **AcadSkil**: language and vocabulary test along with academic completion measure, **FluInt**: fluid intelligence measures, **SelfReg**: working memory and cognitive flexibility measures, **PosAff**: positive affect measures, **NegAff**: negative affect measures, **DisorInt**: internalized disorders, **Somatic**: symptoms, problems or sensory manifestations related to the body, **DisorExt**: dysregulated externalized behaviour.

which could otherwise bias the correlation between phenotypes. The number of categories was derived by a silhouette analysis, that characterized how well separated categories were for different numbers. The silhouette plot identified 4 local maxima of cluster separation, and we used the richest cluster solution, featuring 13 categories (see Figure 13). A first set of categories included behavioural assessments related to negative affect (NegAff), dysregulated externalized behaviour (DisorExt), symptoms, problems or sensory manifestations related to the body (Somatic), and internalized disorder (DisorInt). A second set of categories included physiological measures like blood pressure and body mass index (Physio), as well as reaction time measures (Speed). The third set of categories included items related to self-regulation (SelfReg), working memory and cognitive flexibility (Executives), fluid intelligence groups

(FluInt), language and vocabulary test along with academic completion measure (AcadSkil), and two mixed categories (Mix1, Mix2). The fourth and final category was composed of measures related to positive affect (PosAffect), which was identified robustly using 4, 6 or 13 clusters. Overall, the out-of-scanner behavioural assessments collected by HCP seemed to cluster into 13 easily interpretable categories, and we performed further analyses on the average scores within each category (after normalization of the scores for zero mean and unit variance across subjects).

2.9. Phenotype definitions for heritability and pleiotropy estimates

Brain phenotypes: Subtype weight measures were considered as a brain phenotype since they represented how spatially similar or dissimilar an individual activation map was to the subtype maps. These subtype weights were used as traits to estimate heritability, phenotypic correlation and genetic correlation. Each subject had 5 subtypes weights (Figure 12e) for each task condition (total of 105 brain phenotypic measures per subject).

Out-of-scanner behavioural phenotype: We used the 13 clustered out-of-scanner behavioural measures (Figure 13) to first estimate their heritability, and then to associate with brain phenotypes (subtypes weights) in terms of phenotypic correlation followed by genetic correlation (See supplementary material figure Sup 19 for detailed individual behavioural measures).

In-scanner tasks performance phenotype: In-scanner tasks performance (15 measures) were also used to estimate heritability and to associate with brain phenotypes. For each task condition, we used the in-scanner accuracy measure and the reaction time, except the motor one which does not have accuracy measure. Supplementary Table 1 shows the complete list of in-scanner task performance measures used in this study.

2.10. Univariate analysis of additive genetic variance.

For the heritability estimates, we used the Sequential Oligogenic Linkage Analysis Routines (SOLAR) Eclipse software package [9]). SOLAR uses maximum likelihood variance decomposition methods. The covariance matrix Ω for a pedigree is given by equation (1):

$$\Omega = 2 \cdot \Phi \cdot \sigma_{\mathrm{g}}^2 + \mathrm{I} \cdot \sigma_{\mathrm{e}}^2 \tag{2.1}$$

Where σ_g^2 is the genetic variance due to the additive genetic factors, ϕ is the kinship matrix representing the pairwise kinship coefficients among all individuals, σ_{e}^2 is the variance due to individual — unique environmental effects and measurement error, and I is an identity

[9] http://www.nitrc.org/projects/se_linux

matrix (under the assumption that all environmental effects are uncorrelated among family members). Narrow sense heritability is given by equation (2.2) and is defined as the fraction of phenotypic variance σ_p^2 attributable to additive genetic factors:

$$\mathrm{h}^2 = \sigma_\mathrm{g}^2/\sigma_\mathrm{p}^2 \tag{2.2}$$

The variance parameters are estimated by comparing the observed phenotypic covariance matrix with the covariance matrix predicted by kinship (Almasy and Blangero, 1998) . Significance of the heritability estimate is tested by comparing the likelihood of the model in which σ_g^2 is constrained to zero with that of a model in which σ_g^2 is estimated. Twice the difference between the log-likelihoods of these models yields a test statistic, which is asymptotically distributed as a 1/2:1/2 mixture of χ^2 variables with 1 degree-of-freedom and a point mass at zero. Prior to the heritability estimation, all phenotype (brain and behavioural phenotypes) were adjusted for covariates including sex, age and FD. Inverse Gaussian transformation was also applied to ensure normality of the distribution. Outputs from SOLAR included the heritability estimate (H2r), the significance value (p), and the standard error for each trait (SE).

2.11. Bivariate genetic analyses

We used equation (2.3) to simultaneously quantify the shared genetic variance and the phenotypic correlation between brain phenotypes (weights) and behavioural measures (in and out-of-scanner behavioural measures). To assess this relationship, we used SOLAR, relying on the following model:

$$\rho_p = \sqrt{h_a^2}\sqrt{h_b^2} \cdot \rho_g + \sqrt{1-h_a^2}\sqrt{1-h_b^2} \cdot \rho_e \tag{2.3}$$

Where Pearson's phenotypic correlation ρ_p is decomposed into $\rho_\mathrm{g} and \rho_\mathrm{e}$. ρ_g is the proportion of variability due to shared genetic effects and ρ_e that due to the environment, while h_a^2 and h_b^2 correspond to the previously defined narrow sense heritability for phenotypes a and b, respectively. In our case, one corresponds to the heritability of subtypes weight while the second is the heritability of one of our behavioural scores.

3. Results

3.1. Activation maps for each of the 7 tasks have good face validity are consistent with previous studies

Activation group maps from each task condition showed the expected pattern of activation/deactivation related to each condition, and were visually similar to HCP maps from an introductory paper on tfMRI that were generated on a subset of the sample used here,

and a different preprocessing strategy (Barch et al., 2013). Due to space limitations, we report here only activation maps for the Motor and Language tasks, (see Supplementary Material Figure Sup 20 for all of the 7 tasks). In the Motor task, group functional activation patterns had cortical activation peaks located in the primary sensorimotor cortex, with a gradient of activation along the central sulcus that followed know principles of somatotopic organization: tongue in medio-dorsal position, hand in the area located next to the superior frontal sulcus, and foot in ventral position. Importantly, cortical activation peaks were located in the left hemisphere for right foot or hand movement, and in the right hemisphere, for left foot or hand movement, replicating the known decussation of motor cortico-spinal tract. These observations were all consistent with previous works (Grodd et al., 2001). The language group activation map (contrast of story condition against math condition) revealed a large fronto-temporo-parietal system, mostly left lateralized, with prominent activation in Brodman's areas 44/45 (Broca's area, associated with language production) and Brodman's area 22 (Wernicke's area, associated with language comprehension), as well as the medial temporal cortex and the temporal pole, in accordance with previous descriptions of the language network (Price, 2012). Overall, the group-level activation maps identified distributed brain networks with excellent biological plausibility, and were consistent with previous results from the literature, including those which directly examined HCP data using different analytical choices.

3.2. Individual activation maps cluster in subgroups with similar spatial distribution

We clustered the individual activation maps for each fMRI task condition separately in order to form subgroups of individuals that share similar spatial distribution within the same fMRI task condition. A subtype map was defined as the average activation map of each subgroup (Figure 1d). Figure 14 shows the 5 subtypes for the language and working memory tasks along with the group activation maps, which is the grand mean of all available subjects. For the language task (Figure 14a), the group map revealed a typical fronto-temporo-parietal system, but the subtype maps showed systematic deviations from this pattern. Subtype S1 exhibited stronger activation of fronto-parietal cortices, and underactivation of the bilateral temporal cortices, which mirrored the grand mean. Subtype S2 showed hyper activation of the temporo-parietal junction, along with smaller clusters along the bilateral superior temporal gyri and inferior frontal gyri. Subtype S3 closely resembled S1. Subtype S4 showed a high activation cluster in the superior occipital and superior parietal lobes, while deactivating the main regions of the grand mean. Subtype S5 showed similarity to the grand mean, and was thus composed of subjects who shared a prototypical spatial distribution of activation, yet more strongly activated these regions

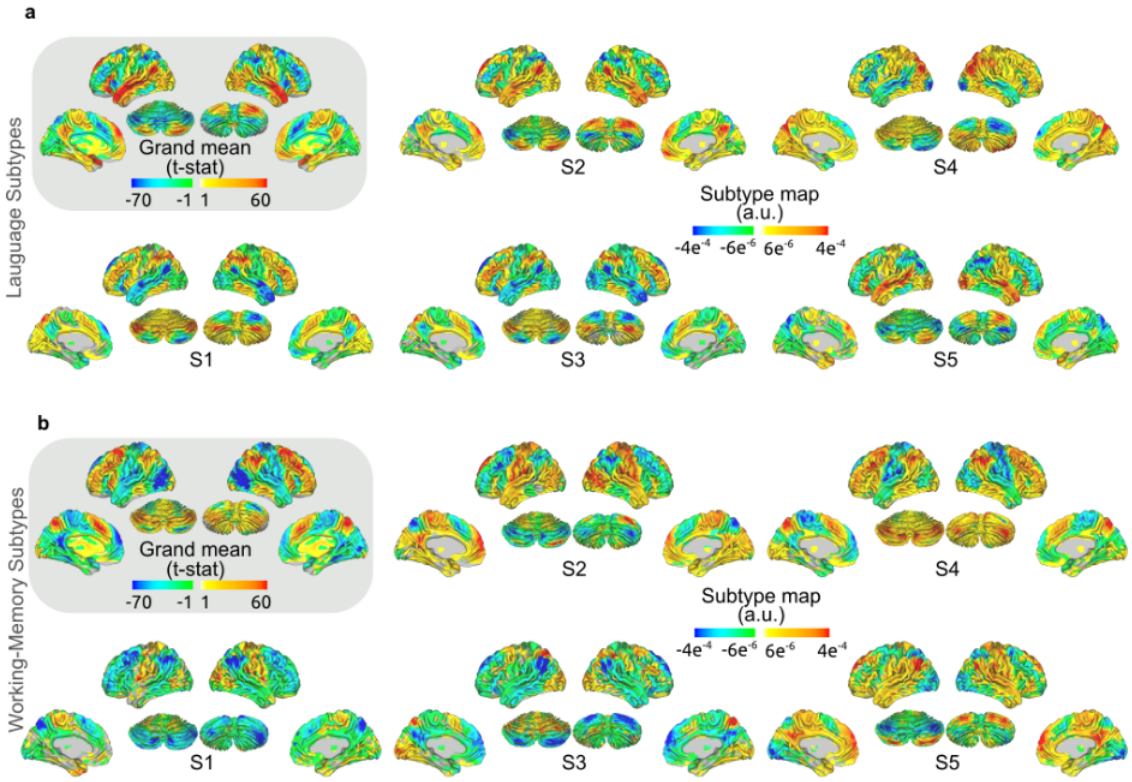


Fig. 14. Subtypes of activation maps - a: Language subtypes - The grand mean for the language statistical parametric maps revealed a cluster of high activation in the superior temporal gyrus during the "Story minus Math" condition. Subtype maps 1 to 5 (S1-5) was defined as the average map of each subgroup or cluster. Subtypes show systematic deviation from the grand mean that range from a predominantly 'deactivating' pattern toward a more 'activating' pattern of brain activity **b: Working memory subtypes** - The grand mean for the working memory statistical parametric maps revealed a cluster of high activation in the shows bilateral activation in the superior frontal gyrus, middle frontal gyrus, inferior frontal gyrus and the inferior parietal lobule during the "0-Back minus 2-Back" condition. Subtype maps 1 to 5 (S1-5) was defined as the average map of each subgroup or cluster. They show activation and deactivation patterns along the superior frontal gyrus (BA 6), middle frontal gyrus (BA 9 and 10), inferior frontal gyrus (BA 47) and the inferior parietal lobule (BA 40).

than other subjects.

For the working memory task (Figure 14b), subtypes showed bilateral activation/deactivation in the superior frontal gyrus (BA 6), middle frontal gyrus (BA 9 and 10), inferior frontal gyrus (BA 47) and the inferior parietal lobule (BA 40). The difference observed within subtypes from the same task are highlighted by the difference of activation or deactivation within these brain regions. For example, subtypes S1 and S2 are mirrored versions of the grand mean activation map of the working memory task, while subtypes S3 and S4 present the same pattern of activation as the grand mean with slight difference in the

intensity. Overall, group activation maps appeared to be an aggregation of heterogeneous activation subgroups. This conclusion generalized to all tasks, and is illustrated in more detail in the supplementary material figure Sup 20 for the remaining 5 tasks.

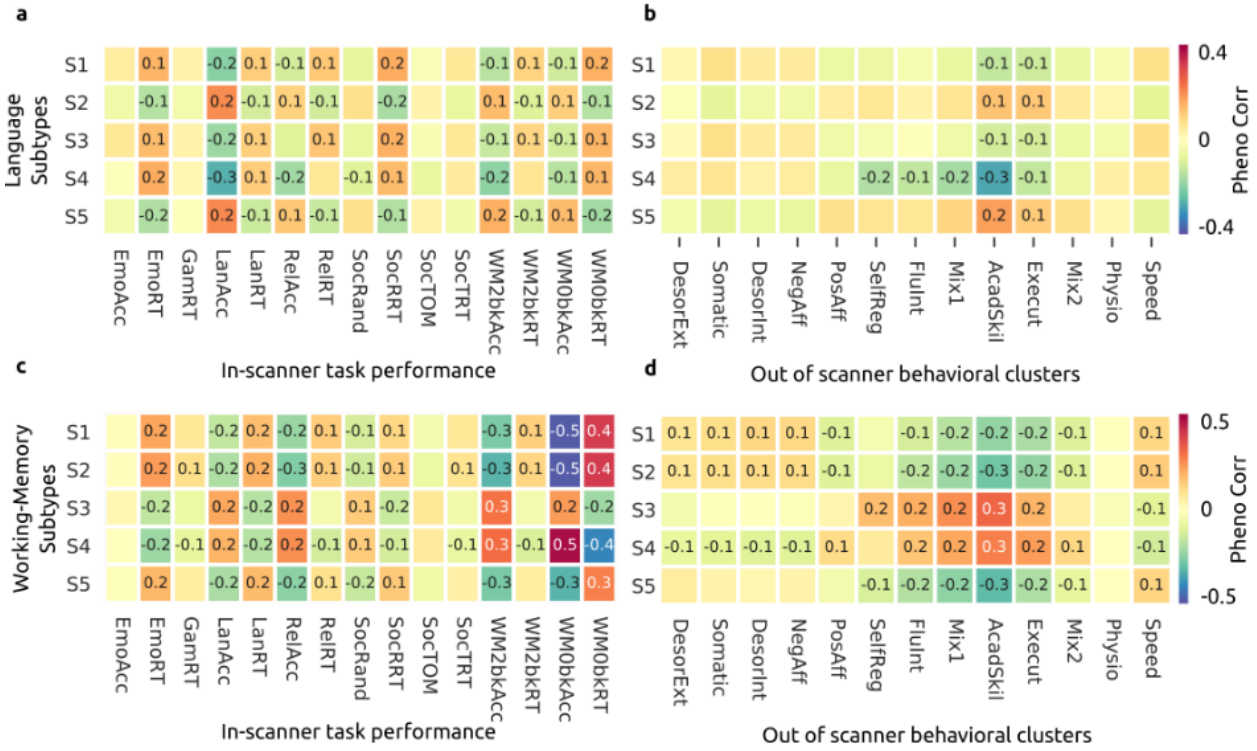


Fig. 15. Phenotypic correlation between brain subtypes and behavioural measures - a: Correlation between language subtypes and in-scanner task performance. The highest correlation estimates are between the language task accuracy (LanAcc) measure and subtypes weights (S1-S5). See supplementary material Table 3 for a list of abbreviations used for the in-scanner task performance. **b - Correlation between language subtypes and out-of-scanner behavioural measures:** The highest correlation estimates are between the academic skills (AcadSkil) behavioural cluster and subtypes weights. See supplementary material figure Sup 19 for a list of abbreviations used for the behavioural clusters. **c - Correlation between working memory subtypes and in-scanner task performance.** The highest correlation estimates are between the working memory task accuracy measures and all subtypes weights. **d - Correlation between working memory subtypes and out-of-scanner behavioural measures.** The highest correlation estimates are between the academic skills (AcadSkil) behavioural cluster and all subtypes weights.

3.2.1. Language and working-memory subtypes weights associate with Behavioural phenotypes. We associated subtypes weights from the 7 tasks with behavioural measures, collected both inside and outside the scanner. Only the weights of the language and the working memory task subtypes showed significant correlations with behavioural phenotypes (See figures Sup 20 and Sup 20 in supplementary material). Notably, subtype weights of both

tasks associated most strongly with performance during these tasks inside the scanner. For example, the language subtype weights showed the highest association with language accuracy performance (figure 15a), and the working memory subtype weights associated most strongly with working memory-related task performance (figure 15c). Subtypes weights were also associated with meaningful outside-of-the-scanner behavioural measures. For example, subtype weights of the language task associated most strongly with the cluster of academic skills (figure 15b), and working memory subtype weights associated most strongly with clusters of fluid intelligence, Mix1, academic skills and executive measures (figure 15d). None of the remaining tasks subtype weights showed significant correlation with their corresponding in-scanner task performance (see supplementary material figures Sup 20 and Sup 20).

3.3. Heritability of subtypes weights ranges from weak to strong

We then estimated the heritability of brain subtypes weights. We first quantified the intersubject spatial similarity (correlation) between individual activation maps for all MZ twins, DZ twins, sibling pairs and unrelated pairs of subjects, separately. The similarity measures ranged from r = 0.05 — 0.25, with a marked variability between tasks. The tasks with the highest inter-subject similarity (social, language, working memory, relational, emotion) showed a clear dosage effect of shared genetics (MZ > DZ > siblings and unrelated, see figure 16a), suggesting genetic control over activation maps. We then tested the heritability of subtypes weights. Significant estimates of additive genetic components were found on subtype weights for all 7 tasks. The heritability strength varied widely across tasks, from close to h2 = 0 — 0.62 (see figure 16b). Some tasks had all associated subtypes highly heritable, e.g. the language task presented the strongest heritability estimates, above h2 = 0.6 for all 5 subtypes, followed by relational, working memory and social tasks. As expected, these activation tasks were also the ones with the highest inter-subject spatial similarity of activation maps and the clearest dosage effects. The gambling task presented the lowest estimate of heritability among all 7 tasks, particularly on the win versus loss contrast (h2 = 0 — 0.18, Figure 5b). The same task also presented the lowest intersubject correlation (r = 0.05 — 0.08, Figure 16a). Overall, some subtypes weights were found to be highly heritable, and the level of heritability was markedly heterogeneous across tasks and across subtypes.

3.4. In-scanner and Out-of-scanner behavioural measures are heritable

Before testing for shared genetic control over brain activation and behaviour, we tested for genetic control over behavioural measures alone. Both in-scanner and out-of-scanner behavioural measures showed significant estimates of additive genetic effects. Regarding

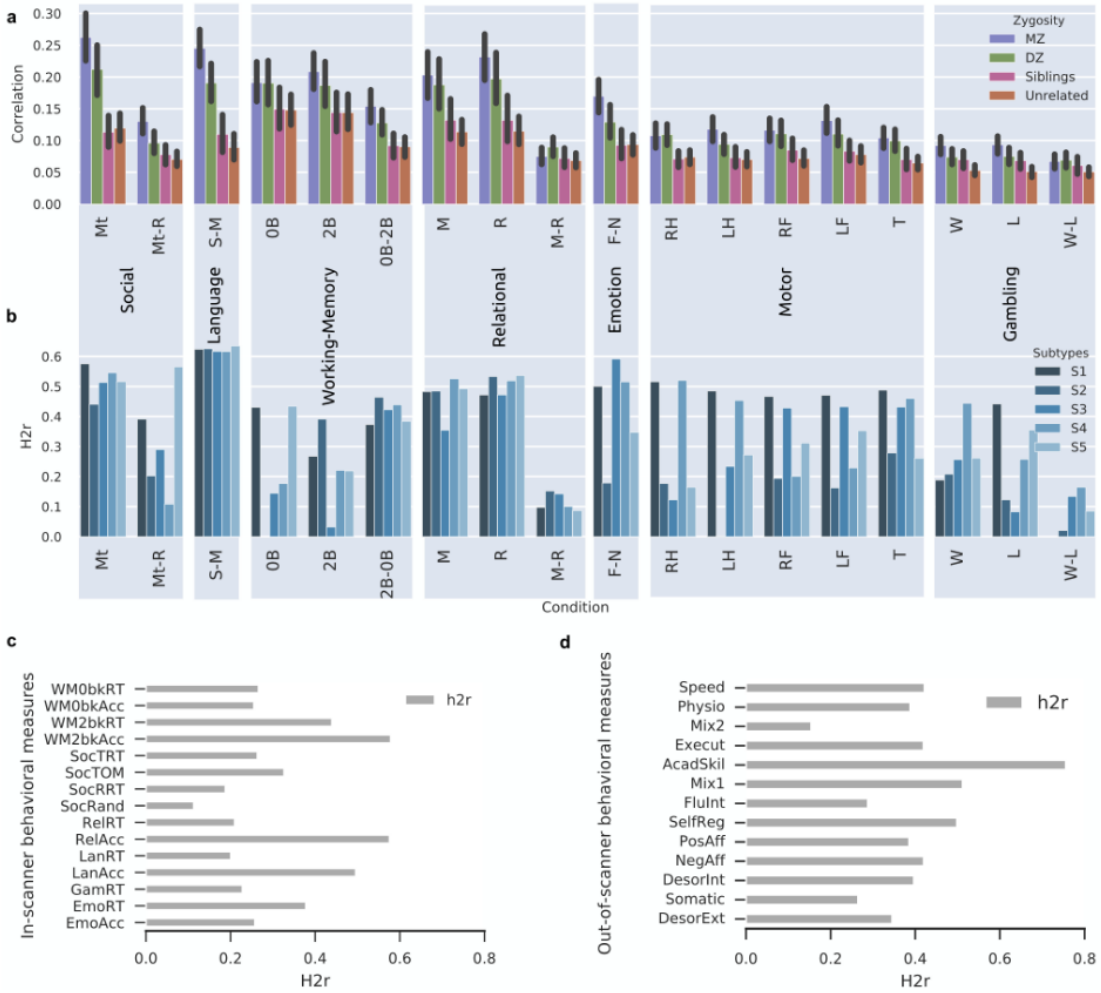


Fig. 16. Heritability estimate - a: Similarity between activation maps by relationship. Average correlation of brain activation maps between pairs of MZ twins, DZ twins, siblings and unrelated individuals. The magnitude of correlations varied substantially across task domains, but a dosage effect was observed in almost every task, with correlation for MZ twins greater than DZ twins, greater than siblings or unrelated individuals. **b: Heritability of subtype weights for all HCP tasks.** Significant heritability was identified for the 7 HCP tasks: gambling task (W: Win, L: Loss, W-L: Win vs Loss), language task (S-M: Story vs Math), motor task (LH: Left Hand, LF: Left Foot, T: Tong, RF: Right Foot, LF: Left Foot), social task (Mt: Mental, Mt-R: Mental vs Random), working memory task (2B: 2-Back, 0B: 0-Back, 2B-0B: 2-Back vs 0-Back). Relational task (M: Match, R: Relational, M-R: Match vs Relational) and the emotional task (F-N: Fear vs Neutral). Each task condition was represented by 5 subtypes' weights. Heritability estimate (H2r) on these subtype weights ranged from 0 to 0.62. **c: Heritability of in-scanner task performance.** Heritability ranged from 0.12 to 0.57 with the highest estimates for working memory and relational accuracy scores. See supplementary material Table 1 for a detailed description of in-scanner behavioural labels. **d: Heritability of out-of-scanner behavioural clusters.** Heritability ranged from 0.15 to 0.75 with the highest estimate for academic skills (AcadSkil). See supplementary material figure Sup 19 for a list of abbreviations used for the clustered behavioural measures.

the in-scanner performance measures, the highest heritability measures were for the relational task (h2RelAcc = 0.57) and working memory task (h2WM2bkAcc=0.57), while the heritability of the remaining measures ranged from h2 = 0.12 — 0.5 (lowest SocRand), see Figure 16c. Among the out-of-scanner behavioural skills, the Academic skills subgroup had the highest heritability estimate (h2AcadSkil= 0.77), followed by Mix1 (h2Mix1= 0.52) and SelfReg (h2SelfReg = 0.50). The remaining subgroups had heritability estimates ranging from h2 = 0.12 — 0.40 (lowest Mix2), see Figure 16d.

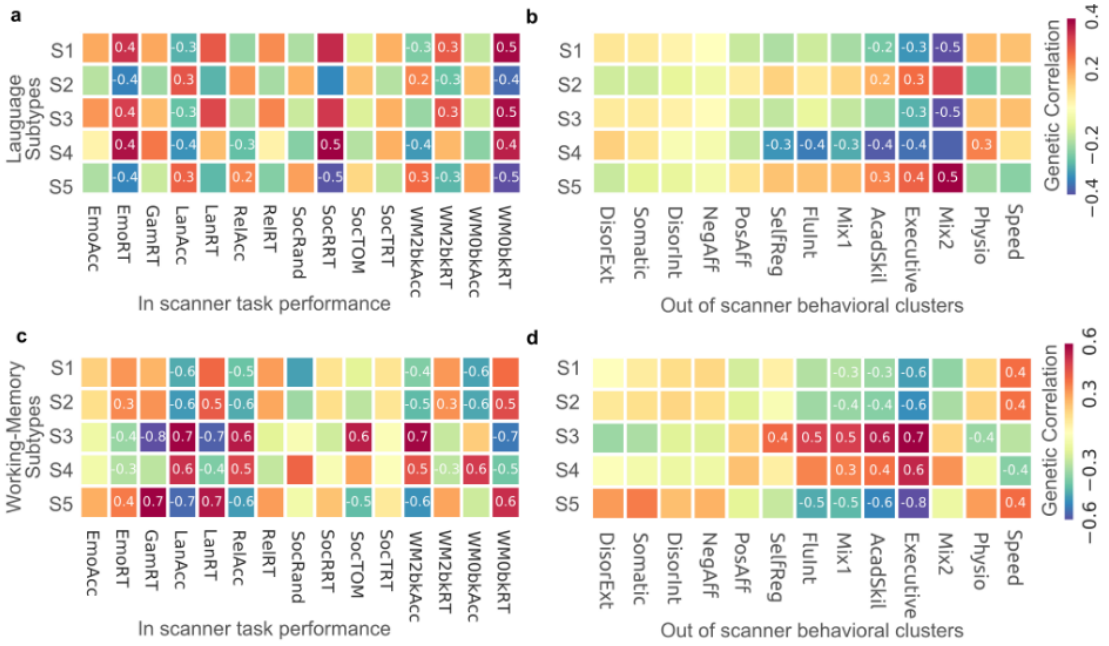


Fig. 17. Genetic correlation between brain phenotypes and behavioural phenotypes. a - Genetic correlation between language subtypes and in-scanner behavioural tests: The highest genetic associations are between the working memory reaction time and the five language subtypes weights. See supplementary material Table 1 for a list of abbreviations used for the in-scanner task performance. **b - Genetic correlation between language subtypes and out-of-scanner behavioural measures.** All five language subtypes weights are genetically correlated with Executive function behavioural cluster, but the highest genetic correlation is between Mix2 and subtypes S1,S3 and S5 (ρ_{gS1} = -0.5, ρ_{gS3}= -0.5 and ρ_{gS5}= 0.5). See supplementary material figure Sup 19 for a list of abbreviations used for the behavioural tests. c - Genetic correlation between working memory subtypes and in-scanner behavioural tests. Working memory subtypes are genetically correlated with all tasks either with reaction time or accuracy measures. Genetic correlation range from -0.8 to 0.7. d - Genetic correlation between working memory subtypes and out-of-scanner behavioural measures: Genetic correlation estimate between working memory subtypes weights and out-of-scanner task performance shows the highest value for Executive function behavioural cluster followed by academic skills.

3.5. Subtypes weights are genetically associated with behaviour

We performed a bivariate genetic analysis to quantify the shared genetic influence (genetic correlation) between subtypes weights for each of the 7 tasks and the behavioural measures inside and outside the scanner. We found significant genetic correlations between the subtypes weights and behavioural measures, especially for the language and working memory tasks (Figure 17), that we report in this section. Other genetic association tests are reported in Supplementary Material Sup 20 and Sup 20.

Figure 17a summarizes the genetic correlation between the language brain activation subtype weights and in scanner task performance. The five language subtypes showed significant shared genetic variance with in-scanner language performance (LanAcc, ρ_g = -0.4 — 0.3), Emotional task reaction time (EmoRT, g = -0.4 — 0.4) and working memory reaction time (WM0bkRT, ρ_g = -0.5 — 0.5). Figure 17b shows the genetic correlation between the language subtype weights and out-of-scanner behavioural clusters. The behavioural measures of executive functions were genetically correlated with all 5 subtype weights (ρ_g = -0.3 — 0.4), and the AcadSkil behavioural measures were genetically correlated with weights of 4 of subtypes (AcadSkil,ρ_g = -0.4 — 0.4). Mix2 (sensory-motor and emotion) showed the highest genetic correlations with subtypes S1, S3 and S5 (ρ_{gS1}= -0.5, ρ_{gS3}= -0.5 and ρ_{gS5}= 0.5), yet this behavioural subtype was not highly heritable (h2 = 0.12). Finally, subtype weight S4 of the language task (figure 17b) shared genetic influences with 6 of the behavioural clusters (SelfReg, FluInt, Mix1, AcadSkil, Executive and Physio).

The five working memory subtypes weights (figure 17c) also showed significant shared genetic variance with in-scanner language (LangAcc,ρ_g= -0.7 — 0.7), relational (RelAcc, ρ_g = -0.6 — 0.6) and working memory 2back (WM2bkAcc, ρ_g = -0.6 — 0.7) tasks performances. Figure 17d shows genetic correlation between the working memory subtype weights and clusters of out-of-scanner behavioural measures. Executive, AcadSkil and Mix1 behavioural clusters were genetically correlated with all 5 WM subtype weights (-0.8 to 0.7). The remaining tasks did not present any significant associations between subtype weights and other behavioural measures inside and outside the scanner.

In summary, we found moderate to strong genetic associations between brain subtypes weights and behavioural measures (in-scanner and out-of-scanner). These associations were particularly strong only for the language and working memory subtypes weights. Subtypes weights from the remaining tasks did not reach significant associations with behavioural measures inside and outside the scanner.

4. Discussion

The goal of this work was to identify subtypes of brain activation and test their phenotypic and genetic association with behavioural measures. Our results demonstrated that subtypes were moderately to highly heritable. Only the language and working memory subtypes associated with behavioural measures, phenotypically and genetically. Our findings emphasize the importance of considering both inter-individual differences and commonalities for the understanding of task-related brain activation, and call into question the validity of using group averages of brain activation patterns.

4.1. Brain subtypes are associated with Behavioural measures

Unsupervised clustering methods are a practical means of characterizing heterogeneous data by delineating homogeneous subpopulations. Subtyping has been predominantly applied to investigate brain subtypes linked to brain disorders but little is known about brain subtypes in a healthy population. Our results on subtypes of brain activation in a healthy sample show that we could distinguish subtypes of similar brain activation with good face validity. Moreover, the weights of these subtypes were shown to be correlated with some behavioural measures. Our results on brain activation subtype are hardly comparable with studies that used group activation maps, but some consistency could be observed in terms of the core regions activated in both studies. For example, in the comprehension language task, our results replicated involvement of bilateral temporal areas in speech comprehension, which has been shown in prior studies (Hickok and Poeppel, 2007). As expected, the superior temporal gyrus/sulcus and medial temporal gyrus were involved in narrative comprehension symmetrically in both hemispheres (AbdulSabur et al., 2014) in the group activation map (see figure 14a grand mean). The 5 subtypes of the language task also presented resemblance with the group average map with marked differences in the combination of regions activated or deactivated that is specific to each subtype. For example, in figure 14a, S1 and S5 are mirrored versions of one another and show opposing association with in-scanner language task performance measures (figure 15a, correlation for S1=-0.2 and S5=0.2).

In this study, only the language and working memory subtypes were associated with a set of behavioural measures. The language subtypes showed the highest correlation with in-scanner performance accuracy for the language task (correlation range from -0.25 to 0.3). Le Guen and colleagues found the same associations between the median activation of the contrast STORY-MATH maps from HCP language task and the behavioural measure inside the scanner (RoP, range from -0.23 to 2.28). They also found correlation with behavioural measures like fluid intelligence, working memory, vocabulary comprehension and reading

decoding (correlation range from -0.25 to 0.25). Their results are consistent with our finding about the correlation on subtypes of language and clustered behavioural measures (figure 14b). Figure 14b shows that language subtypes are associated mostly with AcadSkil clusters (correlation range from -0.3 to 0.2) but less with fluid intelligence cluster (correlation = 0.1). Since our behavioural measures are clustered, it is hard to make a direct comparison with other studies that used HCP behavioural measures to associate with brain activation. Nevertheless, we still find some concordance in the association between some behavioural phenotype and brain phenotypes for language and working memory tasks. For the working memory task, group activation maps (figure 14b, Grand mean) showed typical activation in bilateral dorsal and ventral prefrontal cortex, dorsal parietal cortex and dorsal anterior cingulate. Further, we also see deactivation in the default mode network, including medial prefrontal cortex, posterior cingulate, and the occipital-parietal junction, same as shown in the HCP Consortium paper for task activation (Barch et al., 2013). The working memory subtypes also resemble the grand mean but with notable differences in the combination of activation and deactivation for the above-mentioned regions. As we mentioned above for language subtypes, it is hard to compare our working memory subtypes to other studies since they used group activation maps. Nevertheless, our results for working memory subtypes are consistent with Miller and colleagues (Miller et al., 2009) who used memory retrieval tasks on 14 subjects and looked at intra and inter-individual variability in brain activation patterns. They found that activity patterns of the same individual performing different tasks were more similar than activity patterns of different individuals performing the same task. They also found that individual differences in decision criteria on a recognition test predicted the degree of similarity between any two individuals' patterns of brain activity.

Our working memory subtypes also showed phenotypic association with behavioural measures. In-scanner task performance showed the highest associated with working memory reaction time and accuracy (figure 4c), which is consistent with results from Saliasi (Saliasi et al., 2014) and colleagues who found higher BOLD response in the VLPFC was associated with increased performance accuracy in older adults, in both the baseline and the more complex task condition. Literature on association with working memory brain activation and behaviour, mostly found age-related difference (Yaple et al., 2019). In our study, we cannot test significant age differences since age in the HCP sample ranges from 22 to 35, which is considered a young adult. But we did find associations with behavioural measures outside the scanner. These associations were mainly composed of measures related to cognitive functioning (fluid intelligence, academic skills and executive functions). To our knowledge this is the first study showing association between brain activation from the n-back memory task and behavioural measures outside the scanner. The remaining 5 task activation subtypes

(motor, relational, gambling, emotional and social) did not show any association either with task performance or with any behavioural measures.

4.2. Heritability of brain subtypes and Behavioural measure

Our results showed that most subtypes of brain activation from the 7 tasks had moderate to high genetic influence. The language task showed the highest heritability estimate (mean h2 = 0.52) which is in line with previously reported non-imaging studies on heritability of the language ability (Stromswold, 2001). A meta-analysis of the available evidence concerning a genetic basis for language (Stromswold, 2001) concluded that genetic factors continue to be a predictive element of language ability into adulthood. Brain imaging heritability of language studies likewise indicates a large heritability. We found that superior temporal gyrus and sulcus, inferior frontal gyrus, supplementary motor area, medial superior frontal gyrus, precuneus and middle frontal gyrus were overall highly activated in the 5 subtypes of the language task and these subtypes were heritable. This is in line with previous findings on the heritability of the HCP language task activation maps. Le Guen and colleagues 2018, using the 900 release of HCP, found heritability of language brain activation to range from 0.22 to 0.55 in brain areas overlapping with our study. Abbas Babajani-Feremi ((Babajani-Feremi, 2017) also found high heritability for HCP language tasks (h2 = 0.36). Although our own heritability estimates are higher (h2 = 0.52), we should note that there is a difference on the nature of brain phenotype used (subtypes weights vs voxelwise heritability).

For working memory brain activation subtypes, heritability (0.38 to 0.46) is in line with previous studies showing that a significant proportion of the variance in working memory activation maps may be attributed to genes (Blokland et al., 2011, 2017). These studies from Blokland and colleagues used voxelwise heritability estimate on the region of interest and showed that there is a significant and substantial genetic influence on working memory task-related activation across the brain, with genes accounting for up to 65% of the variance, averaging 33%. For the social, emotional, gambling, motor and relational tasks, we are not aware of any fMRI studies that estimated heritability of activation maps using these same tasks as HCP one. Nevertheless, in the social and gambling domains, some studies explored heritability of task activation fMRI. Van der Meulen and colleagues 2018 showed that children experiencing self-inclusion was associated with activity in anterior cingulate cortex, insula and striatum, but this was not significantly explained by genetics or shared environment. Our results for the social task (mental minus random contrast) showed low to high heritability estimates (0.1 to 0.55), depending on the subtype. This difference between our study and Van Der Meulen results could be attributed to the nature of social tasks used. Our study used a social task designed to capture the neural correlates of theory of

mind, while the other study was interested in social inclusion/exclusion aspects. For the gambling task, Rao and colleagues 2018 investigated the heritability of risk taking and the genetic influence on individual variation in risk-related brain activation. They used a balloon analogue risk task to assess individual risk-taking behaviour in a sample of 244 pairs of young adult twins. They found that there was a moderate heritability (41%) of risk taking. Their results were consistent with only two subtypes (S4 win condition and S1 loose condition, figure 14a), the other subtypes had lower heritability estimates. This difference between their results and our results could be, first, attributed to the difference in tasks used and second, our gambling task showed the lowest intersubject correlation between individual brain maps compared to the other tasks. This lack of correlation may be interpreted as the gambling task used in HCP is less engaging than the other tasks. This phenomenon is largely revealed by the field of neuro-cinematic, where studies use movies as a stimulus for the tfMRI, see (Hasson et al., 2008).

Our results on heritability of behavioural measures are also consistent with the literature. For the out-of-scanner behavioural measures, a previous study (Han and Adolphs, 2019) identified nine latent factors among the HCP behavioural measures (positive social ability, negative affect, general intelligence, self-regulation, attention and processing speed, agreeableness, self-efficacy, language and communication, and competitiveness). The estimated heritability for these nine latent factors ranged from 0.1 to 0.6, using Falconer's formula. Their results are close to our findings, since we also reduced dimensionality of our data and found 13 latent variables (out of 74 measures) and found heritability that ranged from 0.2 to 0.8. In-scanner task performance heritability estimates were also consistent with studies that used the same HCP data (Le Guen et al., 2018; Babajani-Feremi, 2017).

4.3. Genetic correlation between brain subtypes and Behavioural phenotypes

The last goal of this work was to estimate the shared genetic influence between brain phenotypes and behavioural phenotypes. Our results showed significant genetic correlation between brain phenotypes and behavioural phenotypes. Le Guen and colleagues 2018 used HCP data to perform bivariate genetic analyses between the neural activation of the language task and behavioural scores, corresponding to the fMRI task accuracy, fluid intelligence, working memory and language performance. They observed that several parts of the language network along the superior temporal sulcus, as well as the angular gyrus belonging to the math processing network, are significantly genetically correlated with these indicators of cognitive performance. Their findings are consistent with our result of genetic correlation between brain activation and behavioural measures. We both found significant association

between language brain phenotypes and in scanner language task performance accuracy, and with cognitive abilities. In addition we found that some of our language subtypes were also associated with task performances and reaction time from other non-language tasks, namely the working memory, emotional social and the relational tasks. The same genetic association also happened between working memory brain phenotypes and behavioural ones. The observed genetic correlation between brain and behavioural phenotypes in this work shed light on the genetic root of individual variability. Subtypes are shown to have shared genetic aetiologie with behavioural measures. Our results are a step forward to understand the nature of inter-individual brain variability in relation to behavioural manifestation and their genetic link.

4.4. Cognitive factors underlying brain subtypes?

There are many possible factors that could be contributing to individual heterogeneity in brain activity, including individual differences in cognitive processing, psychological states or traits, physiology, anatomy, personality, and genetics. Kirchhoff and colleagues 2006 tested the neural correlates of different memory encoding strategies and found that self-reported use of strategies did capture significant variation in memory performance and regional brain activity. This suggests that different cognitive strategies used by the individuals to execute a given task may be reflected by homogeneous subgroups of activation patterns. Our finding on the correlation between subtypes and behavioural measures supports that hypothesis, in which brain phenotype (only for the language and working memory task) associated with the task performance of the related task. For example, in the language task (figure 15a), subjects who exhibited spatial brain activation patterns similar to the subtype activation map S1 have positively correlated behavioural scores of the in-scanner language task, while subjects who exhibited brain activation patterns similar to subtype 2 (S2), have negatively correlated behavioural scores for the in-scanner behavioural language task. We can speculate that subjects associated with subtype 1 have used different cognitive strategies to perform the language task compared to subjects associated with subtype 2. This strategy shifting and differences in normal population could be extended to brain disease studies. For example, is it possible that certain subtypes are more at risk for a specific brain disorder and are indeed using a less efficient cognitive strategy to get the task done. More direct characterization of cognitive strategies employed during tasks will be needed to shed more light on the cognitive and physiological basis of interindividual heterogeneity in brain activation.

4.5. Limitations

This study has some limitations. First, in this work we made the choice of preprocessing the HCP data and not using the HCP preprocessed data mainly to implement motion

censoring. Since head motion is highly heritable, we opted for using the scrubbing method to ensure that time points with excessive motion are removed. Figure Sup 18 in the supplementary material shows the decreased heritability estimate of head motion after applying scrubbing. Despite our efforts to reduce the effects of head motion, we still need to be careful in the interpretation of our result, since some effects of head motion may remain in the data and bias heritability estimate (Bolton et al., 2020).

Second, in this work, we could not accurately estimate the effect of the shared environment between siblings, since HCP did not provide the household information necessary to model it. We opted for using the mother ID as a household ID. This is not ideal, since split families (ex: divorced) are still considered to share the same familial environments, which is not necessarily the case. Our heritability estimates could be slightly overestimated for that reason.

Third, we should caution about the possibly limited generalizability of our study since the majority of participants were Caucasian (European origin). HCP did offer information about the ethnicity of the participants but this information was not sufficient to use as confounding elements in our analysis.

4.6. Future directions

The recent efforts in associating subpopulation of brain disorders and brain patterns has led to a shift in the way we conceive a disease. These studies helped clarify the marked heterogeneity that exists in psychiatric illnesses by associating structural and functional subtypes with different clinical diagnostics for the same disorder. Studies showed significant associations between brain subtypes from structural measures and clinical diagnosis, but they lacked the explanation behind these associations. Our work on healthy populations could explain better the nature of those subgroups in the normal population by pointing to the biological underpinning of their manifestation, and their relationship with behaviour. As a future direction, one could compare the distribution of the subtypes of activation maps with gene expression in the Allen brain atlas (Sunkin et al., 2013). A work by Grasby 2020 and colleagues explored a similar relationship by associating between cortical thickness and neurological, psychological, and behavioural traits. The highly polygenic architecture of the cortex suggests that distinct genes are involved in the development of specific cortical areas. Moreover, they found evidence that brain structure is a key phenotype along the causal pathway that leads from genetic variation to differences in general cognitive function. Similar work could be performed by focusing on the functional organization of brain maps from subtypes. The results will help make sense of the relation between subtypes of brain

activation, genetic and behavioural measure in our work. This will guide other studies on brain disorder, and what variables are worth looking for to explain subgroups of the same disease and their relation to behavioural phenotypes.

Supplementary material

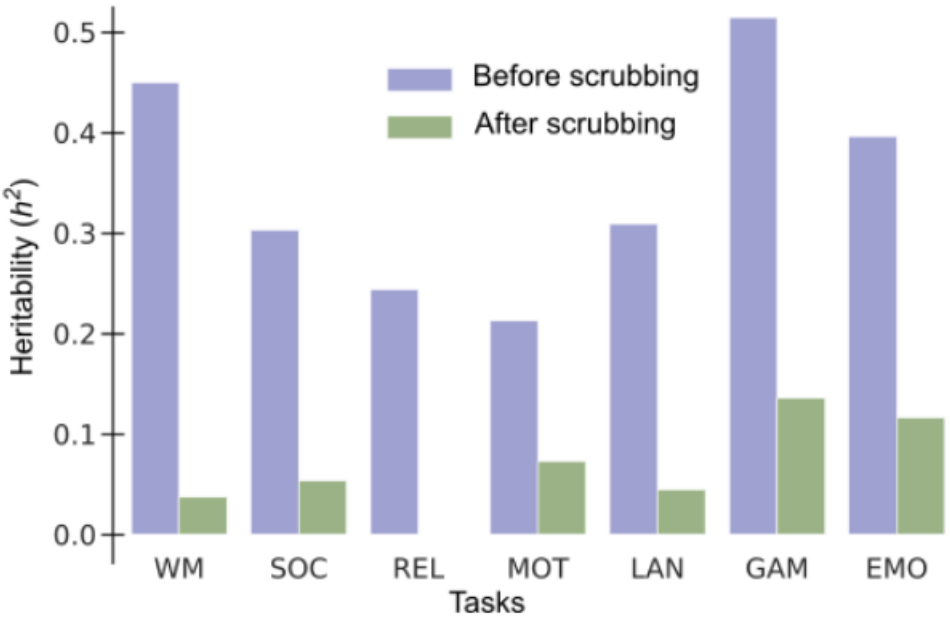


Fig. Sup 18. Heritability of framewise displacement (FD) measure before and after scrubbing on the 7 tasks. WM: working memory, SOC: social, REL: relational, MOT: motor, LAN: language, GAM: gambling and EMO: emotional. Heritability of FD was estimated before and after scrubbing to shows how motion reduction used in this paper is efficient in reducing the heritability estimate that is known to affect further estimate of heritability on brain phenotypes.

columnHeader	assessment	fullDisplayName
EmoAcc	Emotion	OVERALL Emotion Task accuracy
EmoRT	Emotion	OVERALL Emotion Task Reaction Time
GamRT	Gambling	Gambling Task Overall Reaction Time 'Larger'
LanAcc	Language	Language Task OVERALL accuracy
LanRT	Language	Language Task OVERALL median Reaction Time
RelAcc	Relational	Relational Task OVERALL accuracy
RelRT	Relational	Relational Task OVERALL Reaction Time
SocRand	Social	Social Task Percentage 'Random' in Random condition
SocRRT	Social	Social Task Median Reaction Time 'Random' in Random condition
SocTOM	Social	Social Task Percentage 'TOM' in Social (TOM) condition
SocTRT	Social	Social Task Median Reaction Time 'TOM' in Social (TOM) condition
WM2bkAcc	Working Memory	Working Memory Task Accuracy for 2-back
WM2bkRT	Working Memory	Working Memory Task Median Reaction Time for 2-back
WM0bkAcc	Working Memory	Working Memory Task Accuracy for 0-back
WM0bkRT	Working Memory	Working Memory Task Median Reaction Time for 0-back

Table 3. In-scanner task performance description - In-scanner accuracy and reaction time measures used in this study.


Speed
Physio
Mix2
Executive
AcadSkil
Mix1
FluInt
SelfReg
PosAff
NegAff
DesorInt
Somatic
DesorExt
4
2
0
ER40_CRT Emotion Recognition Penn Emotion Recognition Test: Correct Responses Median
IWRD_RTC Verbal Episodic Memory Penn Word Memory Test: Median Reaction Time for Correct Responses
VSPLOT_CRTE Spatial Orientation Penn Line Orientation: Median Reaction Time
SCPT_TPRT Sustained Attention Short Penn Continuous Performance Test: Median Response Time
BPDiastolic Physical Health Diastolic blood pressure of subject
BPSystolic Physical Health Systolic blood pressure of subject
BMI Physical Health Body Mass Index
Strength_Unadj Grip Strength Dynamometry NIH Toolbox Grip Strength Test: Unadjusted Scale Score
Mars_Final Contrast Sensitivity Mars Final Contrast Sensitivity Score
Taste_Unadj NIH Toolbox Regional Taste Intensity Age 12+ Unadjusted Scale Score
GaitSpeed_Comp Locomotion NIH Toolbox 4-Meter Walk Gait Speed Test: Computed Score
ER40_CR Penn Emotion Recognition Test: Number of Correct Responses
Flanker_Unadj Executive Function NIH Toolbox Flanker Inhibitory Control and Attention Test: Unadjusted Scale
CardSort_Unadj Executive Function NIH Toolbox Dimensional Change Card Sort Test: Unadjusted Scale Score
ProcSpeed_Unadj Processing Speed NIH Toolbox Pattern Comparison Processing Speed Test: Unadjusted Scale
PicVocab_Unadj Language/Vocabulary Comprehension NIH Toolbox Picture Vocabulary: Unadjusted Scale Score
SSAGA_Educ Demographics Years of education completed
NEOFAC_O Personality Five Factor Model NEO-FFI Openness to Experience
VSPLOT_TC Spatial Orientation Penn Line Orientation: Total Number Correct
ReadEng_Unadj Language/Reading Decoding NIH Toolbox Oral Reading Recognition Test: Unadjusted Scale Score
Endurance_Unadj Endurance NIH Toolbox 2-minute Walk Endurance Test: Unadjusted Scale Score
IWRD_TOT Verbal Episodic Memory Penn Word Memory Test: Total Number of Correct Responses
PicSeq_Unadj Episodic Memory NIH Toolbox Picture Sequence Memory Test: Unadjusted Scale Score
Dexterity_Unadj Dexterity NIH Toolbox 9-hole Pegboard Dexterity Test: Unadjusted Scale Score
SCPT_TN Sustained Attention Short Penn Continuous Performance Test: True Negatives
SCPT_TP Sustained Attention Short Penn Continuous Performance Test: True Positives
PMAT24_A_RTCR Fluid Intelligence Penn Progressive Matrices: Median Reaction Time for Correct Responses
PMAT24_A_CR Fluid Intelligence Penn Progressive Matrices: Number of Correct Responses
DDisc_AUC_40K Self-regulation/Impulsivity Delay Discounting: Area Under the Curve for Discounting of $40,00
DDisc_AUC_200 Self-regulation/Impulsivity Delay Discounting: Area Under the Curve for Discounting of $200
NEOFAC_E Personality Five Factor Model NEO-FFI Extroversion
NEOFAC_A Personality Five Factor Model NEO-FFI Agreeableness
EmotSupp_Unadj Social Relationships NIH Toolbox Emotional Support Survey: Unadjusted Scale Score
Friendship_Unadj Social Relationships NIH Toolbox Friendship Survey: Unadjusted Scale Score
InstruSupp_Unadj Social Relationships NIH Toolbox Instrumental Support Survey: Unadjusted Scale Score
NEOFAC_C Personality Five Factor Model NEO-FFI Conscientiousness
SelfEff_Unadj Stress and Self Efficacy NIH Toolbox Self-Efficacy Survey: Unadjusted Scale Score
MeanPurp_Unadj Psychological Well-being NIH Toolbox Meaning and Purpose Survey: Unadjusted Scale Score
LifeSatisf_Unadj Psychological Well-being NIH Toolbox General Life Satisfaction Survey: Unadjusted Scale Score
PosAffect_Unadj Psychological Well-being NIH Toolbox Positive Affect Survey: Unadjusted Scale Score
NEOFAC_N Personality Five Factor Model NEO-FFI Neuroticism
PercStress_Unadj Stress and Self Efficacy NIH Toolbox Perceived Stress Survey: Unadjusted Scale Score
AngHostil_Unadj Negative Affect NIH Toolbox Anger-Hostility Survey: Unadjusted Scale Score
FearAffect_Unadj Negative Affect NIH Toolbox Fear-Affect Survey: Unadjusted Scale Score
Sadness_Unadj Negative Affect NIH Toolbox Sadness Survey: Unadjusted Scale Score
AngAffect_Unadj Negative Affect NIH Toolbox Anger-Physical Aggression Survey: Unadjusted Scale Score
PercReject_Unadj Social Relationships NIH Toolbox Perceived Rejection Survey: Unadjusted Scale Score
Loneliness_Unadj Social Relationships NIH Toolbox Loneliness Survey: Unadjusted Scale Score
PercHostil_Unadj Social Relationships NIH Toolbox Perceived Hostility Survey: Unadjusted Scale Score
ASR_Intn_T Life Function ASR Intrusive Raw Score
ASR_Anxd_Pct Life Function ASR Anxious/Depressed Raw Score
DSM_Anxi_Pct Life Function ASR DSM Anxiety Problems Raw Score
DSM_Depr_Pct Life Function ASR DSM Depressive Problems Raw Score
ASR_Thot_Pct Life Function ASR Thought Problems Raw Score
DSM_Avoid_Pct Life Function ASR DSM Avoidant Personality Problems Raw Score
ASR_Witd_Pct Life Function ASR Withdrawn Raw Score
DSM_Adh_Pct Life Function ASR DSM AD/H Problems Raw Score
ASR_Attn_Pct Life Function ASR Attention Problems Raw Score
PainInterf_Tscore Pain Intensity and Interference Surveys NIH Toolbox Pain Interference Survey Age 18+: T-score
FearSomat_Unadj Negative Affect NIH Toolbox Fear-Somatic Arousal Survey: Unadjusted Scale Score
DSM_Somp_Pct Life Function ASR DSM Somatic Problems Raw Score
ASR_Soma_Pct Life Function ASR Somatic Complaints Raw Score
PSQI_Score Pittsburgh Sleep Quality Index
DSM_Antis_Pct Life Function ASR DSM Antisocial Personality Problems Raw Score
ASR_Rule_Pct Life Function ASR Rule Breaking Behavior Raw Score
ASR_Extn_T Life Function ASR Externalizing Raw Score
ASR_Aggr_Pct Life Function ASR Aggressive Behavior Raw Score
ASR_Intr_Pct Life Function ASR Intrusive Raw Score
AngAggr_Unadj Negative Affect NIH Toolbox Anger-Physical Aggression Survey: Unadjusted Scale Score

Fig. Sup 19. Behavioural measures clusters. Group activation maps for motor and language tasks - 74 behavioural test are clustered using k-means. The dendrogram show 13 clusters cut off that are Speed: task reaction time measures, Physio: physiological measures, Mix2 and Mix1: a mixed categories of measures, Executive: working memory and cognitive flexibility measures, AcadSkil: language and vocabulary test along with academic completion measure, FluInt: fluid intelligence measures, SelfReg: working memory and cognitive flexibility measures, PosAff: positive affect measures, NegAff: negative affect measures, DisorInt: internalized disorders, Somatic: symptoms, problems or sensory manifestation related to the body, DisorExt: dysregulated externalized behaviour.

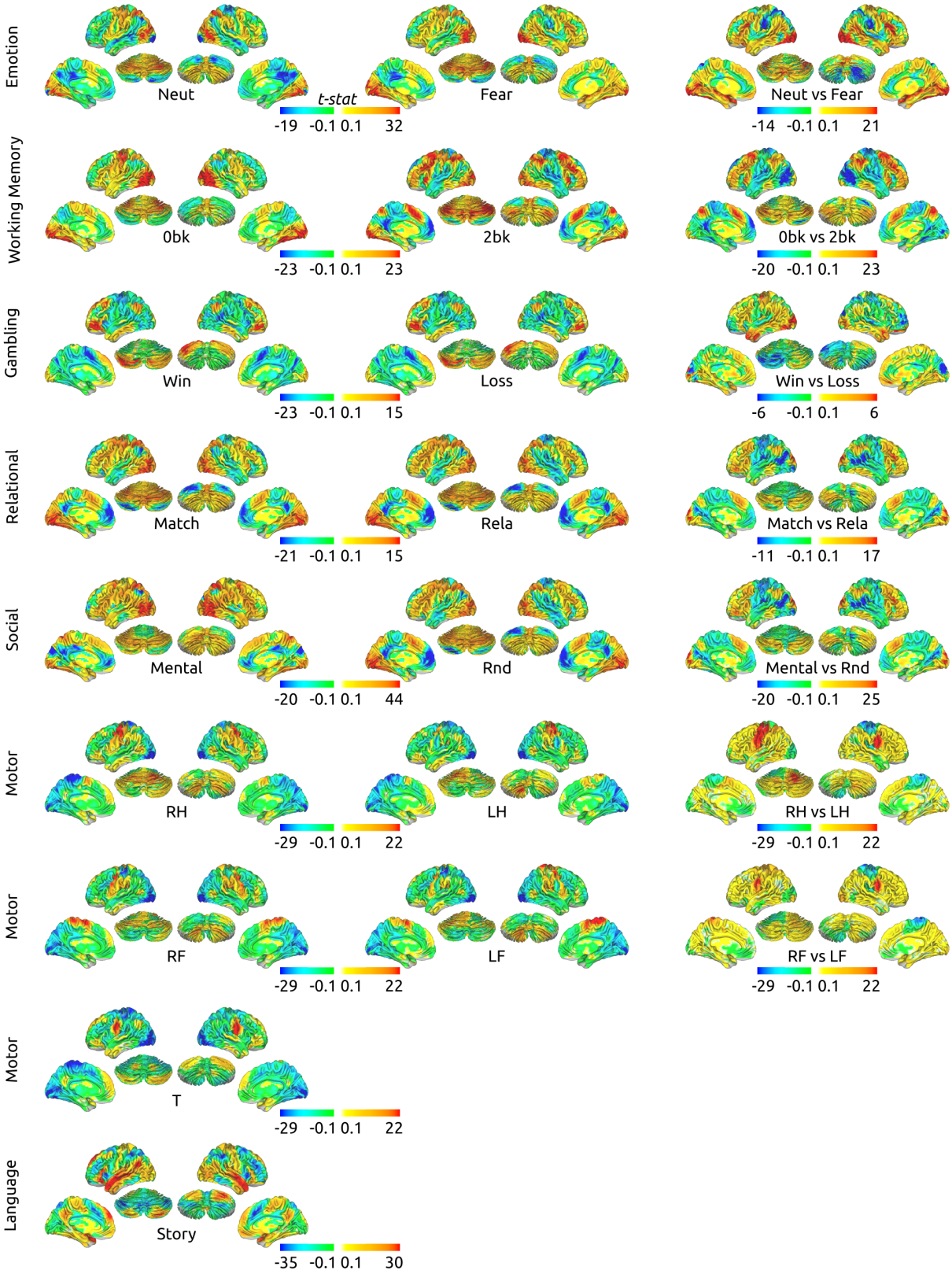
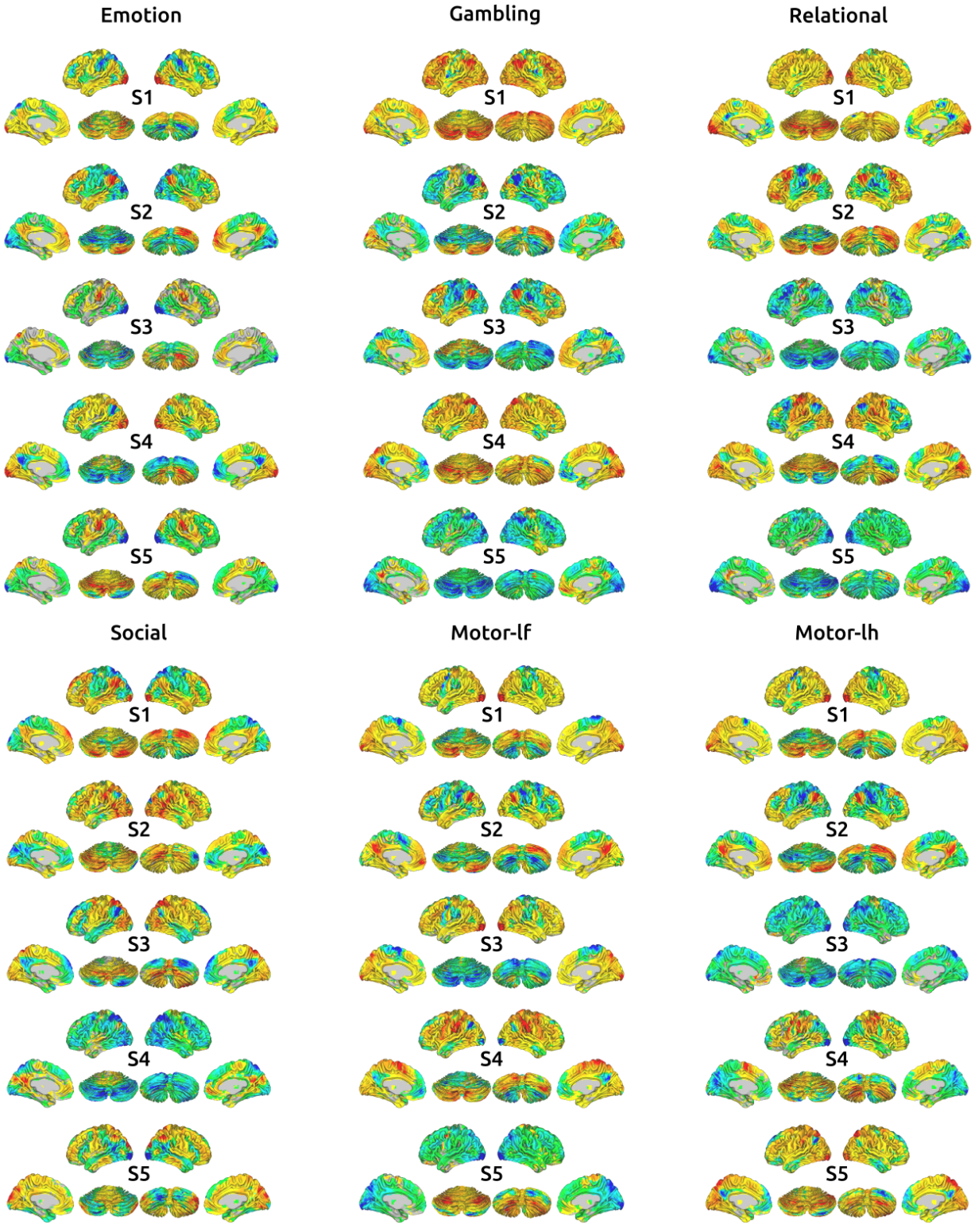

Emotion
Neut
t-stat
-19 -0.1 0.1 32
Fear
Neut vs Fear
-14 -0.1 0.1 21
Working Memory
0bk
-23 -0.1 0.1 23
2bk
0bk vs 2bk
-20 -0.1 0.1 23
Gambling
Win
-23 -0.1 0.1 15
Loss
Win vs Loss
-6 -0.1 0.1 6
Relational
Match
-21 -0.1 0.1 15
Rela
Match vs Rela
-11 -0.1 0.1 17
Social
Mental
-20 -0.1 0.1 44
Rnd
Mental vs Rnd
-20 -0.1 0.1 25
Motor
RH
-29 -0.1 0.1 22
LH
RH vs LH
-29 -0.1 0.1 22
Motor
RF
-29 -0.1 0.1 22
LF
RF vs LF
-29 -0.1 0.1 22
Motor
T
-29 -0.1 0.1 22
Language
Story
-35 -0.1 0.1 30

Fig. Sup 20. Group activation maps for the 7 HCP tasks - Emotion task: Group activation maps for neutral condition (Neut), fear condition (Fear) and the contrast Neut vs Fear. Working memory task: Group activation maps for 0-back (0bk), 2-back (2bk) and the contrast 0bk vs 2bk. Gambling task: Group activation maps for win (Win), loss (Loss) and the contrast Win vs Loss. Relational task: Group activation maps for Match, relational (Rel) and the contrast Match vs Rel. Social task: Group activation maps for Mental, random (Rnd) and the contrast Mental vs Rnd. Motor task: Group activation maps for left hand (LH), right hand (RH), left foot (LF), right foot (RF), tongue (T), contrast RF vs LF and contrast RH vs LH. Language task: Group activation maps for story.


Emotion
S1
S2
S3
S4
S5
Gambling
S1
S2
S3
S4
S5
Relational
S1
S2
S3
S4
S5
Social
S1
S2
S3
S4
S5
Motor-lf
S1
S2
S3
S4
S5
Motor-lh
S1
S2
S3
S4
S5

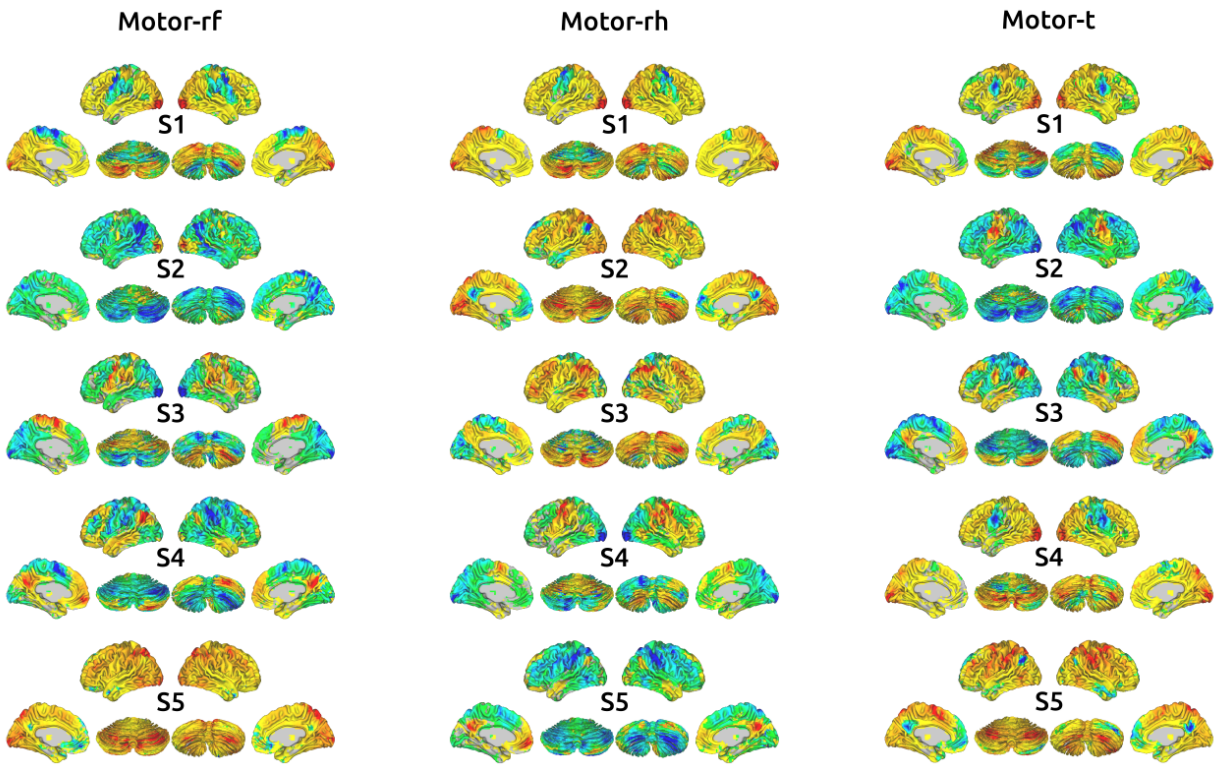
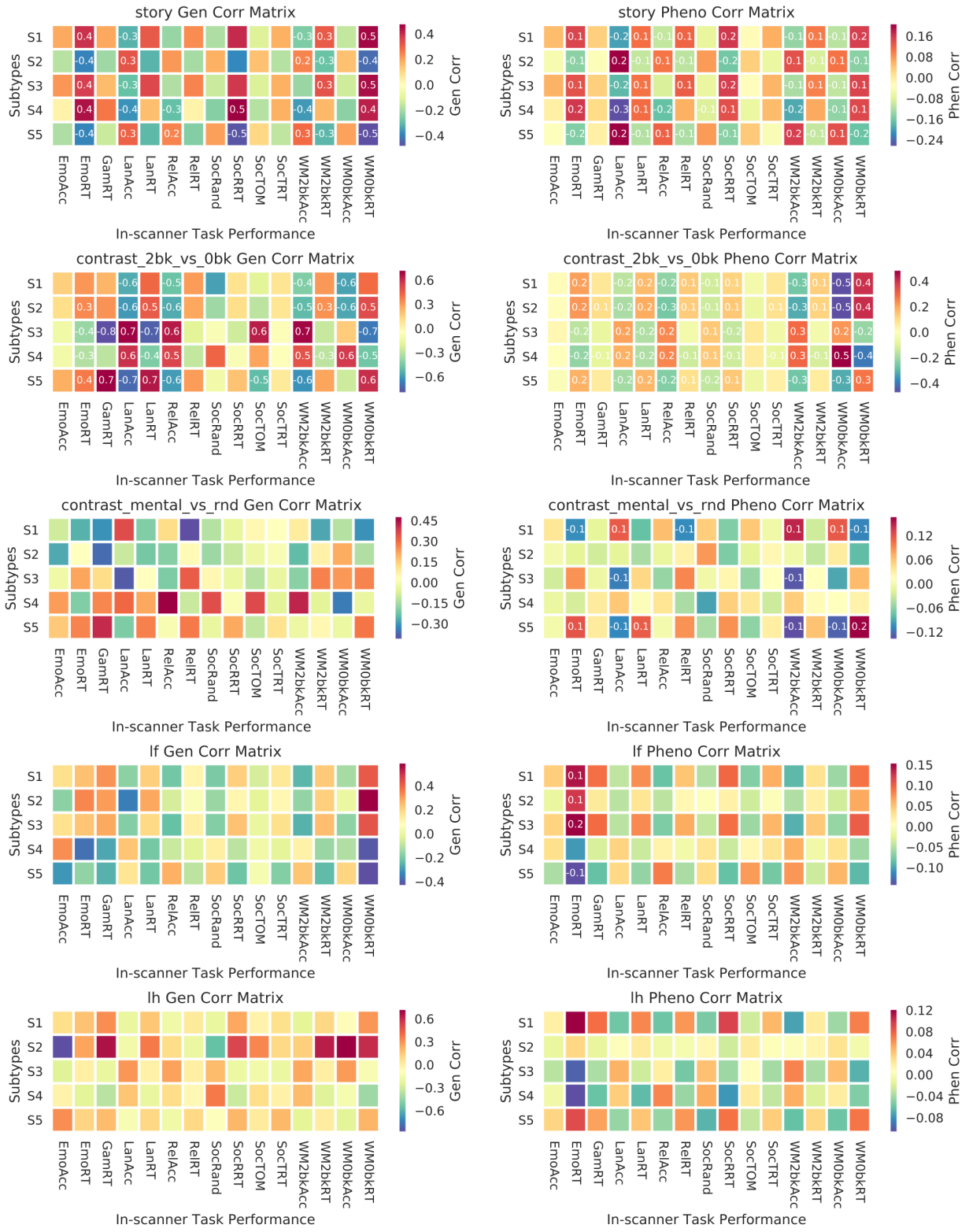
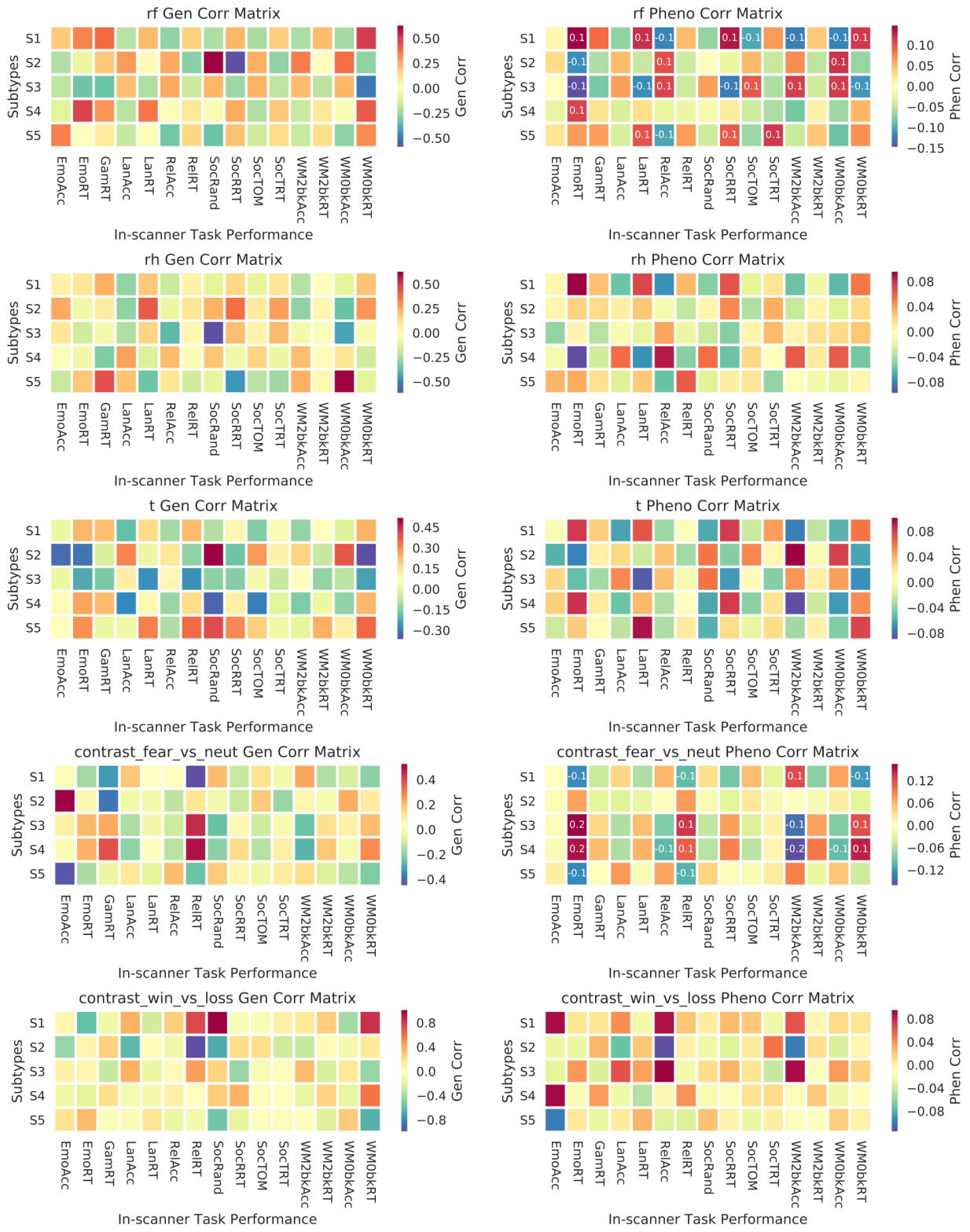


Fig. Sup 20. Task activation subtypes: Fives subtypes maps (S1 to S5) for each of the social, motor, emotion, gambling and relational tasks. Working memory and language subtypes are shown on the manuscript figure 14


story Gen Corr Matrix
story Pheno Corr Matrix
contrast_2bk_vs_0bk Gen Corr Matrix
contrast_2bk_vs_0bk Pheno Corr Matrix
contrast_mental_vs_rnd Gen Corr Matrix
contrast_mental_vs_rnd Pheno Corr Matrix
If Gen Corr Matrix
If Pheno Corr Matrix
Ih Gen Corr Matrix
Ih Pheno Corr Matrix
Subtypes
S1
S2
S3
S4
S5
EmoAcc
EmoRT
GamRT
LanAcc
LanRT
RelAcc
RelRT
SocRand
SocRRT
SocTOM
SocTRT
WM2bkAcc
WM2bkRT
WM0bkAcc
WM0bkRT
In-scanner Task Performance
Gen Corr
Phen Corr


rf Gen Corr Matrix
rf Pheno Corr Matrix
rh Gen Corr Matrix
rh Pheno Corr Matrix
t Gen Corr Matrix
t Pheno Corr Matrix
contrast_fear_vs_neut Gen Corr Matrix
contrast_fear_vs_neut Pheno Corr Matrix
contrast_win_vs_loss Gen Corr Matrix
contrast_win_vs_loss Pheno Corr Matrix
Subtypes
S1
S2
S3
S4
S5
EmoAcc
EmoRT
GamRT
LanAcc
LanRT
RelAcc
RelRT
SocRand
SocRRT
SocTOM
SocTRT
WM2bkAcc
WM2bkRT
WM0bkAcc
WM0bkRT
In-scanner Task Performance
Gen Corr
Phen Corr

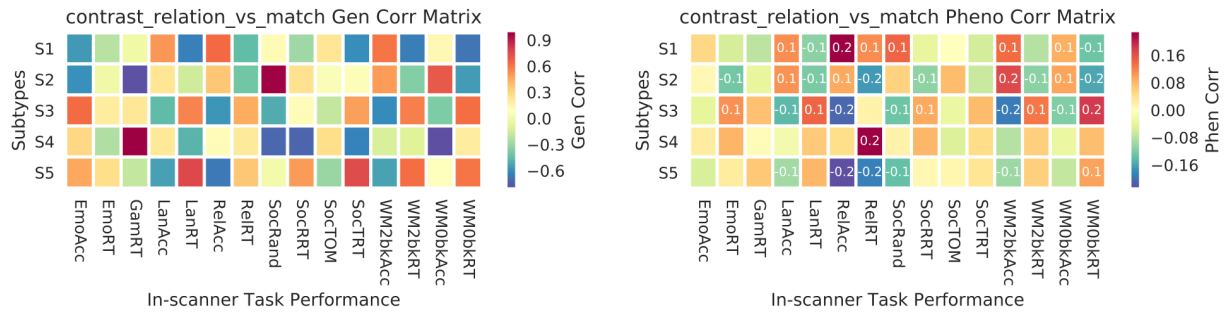
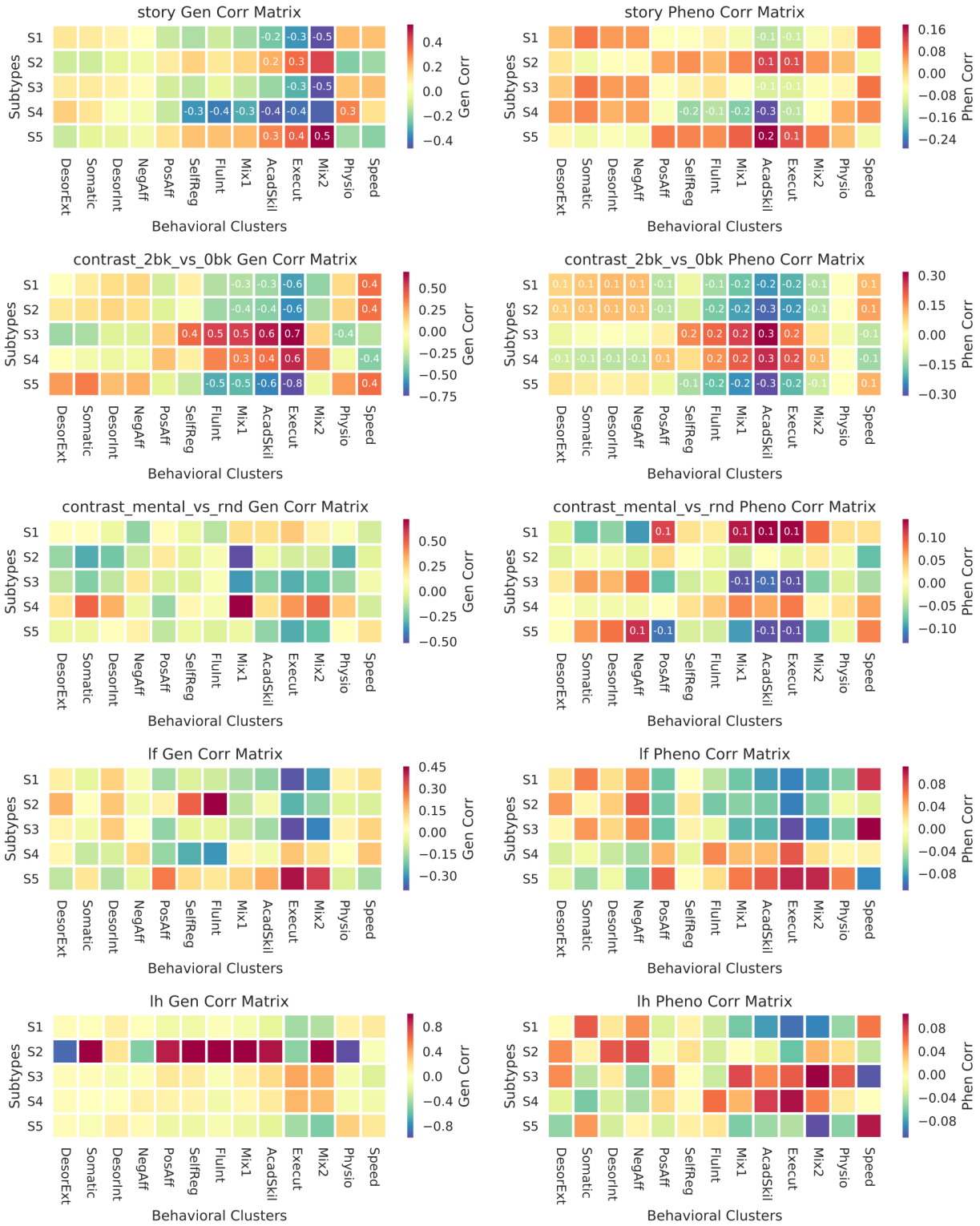
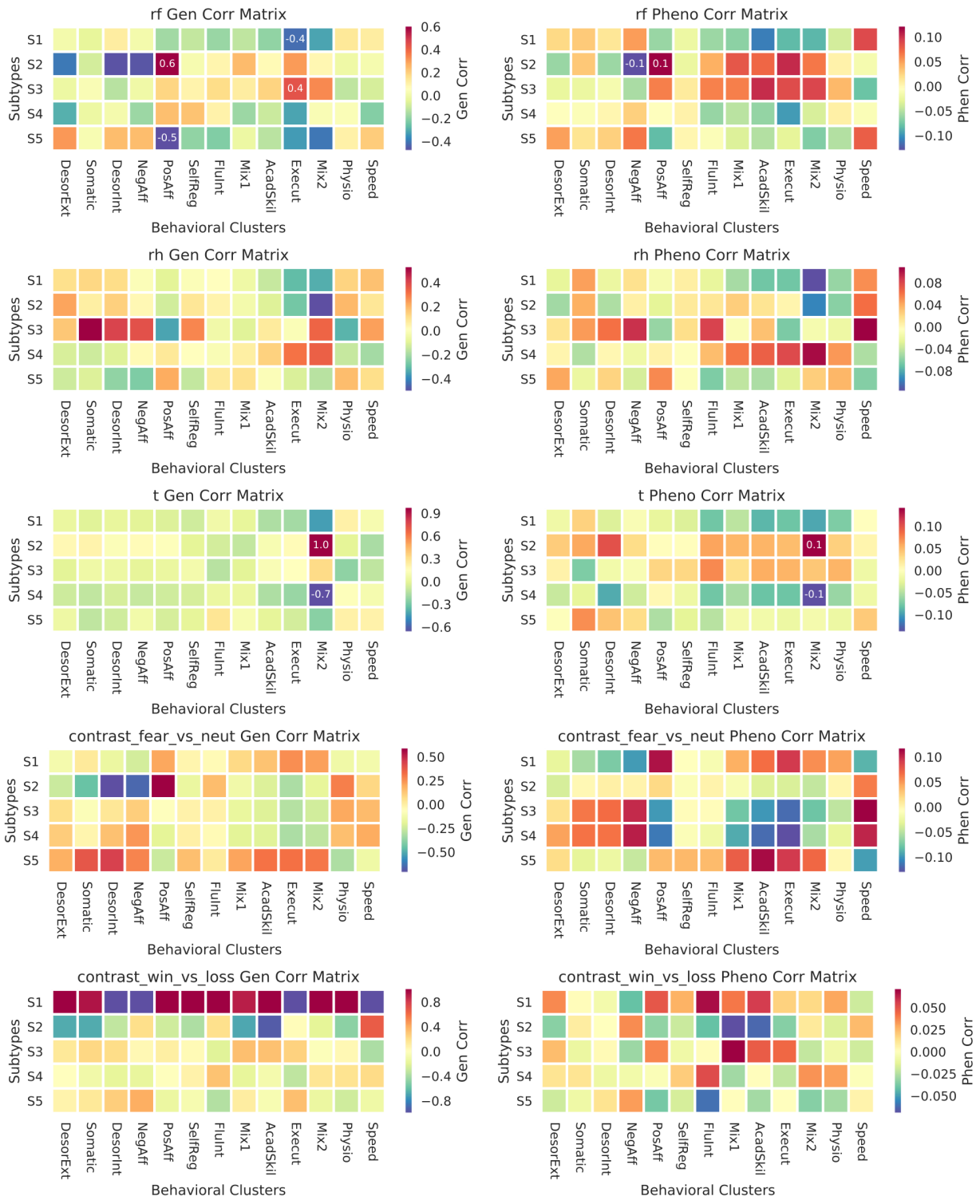


Fig. Sup 20. Genetic and phenotypic correlation for in-scanner behavioural performance: Left column represents genetic correlation between in scanner behavioural performance and brain subtypes weights. Right column represents phenotypic correlation between in scanner behavioural performance and brain subtypes weights.


story Gen Corr Matrix
story Pheno Corr Matrix
contrast_2bk_vs_0bk Gen Corr Matrix
contrast_2bk_vs_0bk Pheno Corr Matrix
contrast_mental_vs_rnd Gen Corr Matrix
contrast_mental_vs_rnd Pheno Corr Matrix
lf Gen Corr Matrix
lf Pheno Corr Matrix
lh Gen Corr Matrix
lh Pheno Corr Matrix
Subtypes
S1
S2
S3
S4
S5
DesorExt
Somatic
DesorInt
NegAff
PosAff
SelfReg
FluInt
Mix1
AcadSkil
Execut
Mix2
Physio
Speed
Behavioral Clusters
Gen Corr
Phen Corr


rf Gen Corr Matrix
rf Pheno Corr Matrix
rh Gen Corr Matrix
rh Pheno Corr Matrix
t Gen Corr Matrix
t Pheno Corr Matrix
contrast_fear_vs_neut Gen Corr Matrix
contrast_fear_vs_neut Pheno Corr Matrix
contrast_win_vs_loss Gen Corr Matrix
contrast_win_vs_loss Pheno Corr Matrix
Subtypes
S1
S2
S3
S4
S5
DesorExt
Somatic
DesorInt
NegAff
PosAff
SelfReg
FluInt
Mix1
AcadSkil
Execut
Mix2
Physio
Speed
Behavioral Clusters
Gen Corr
Phen Corr

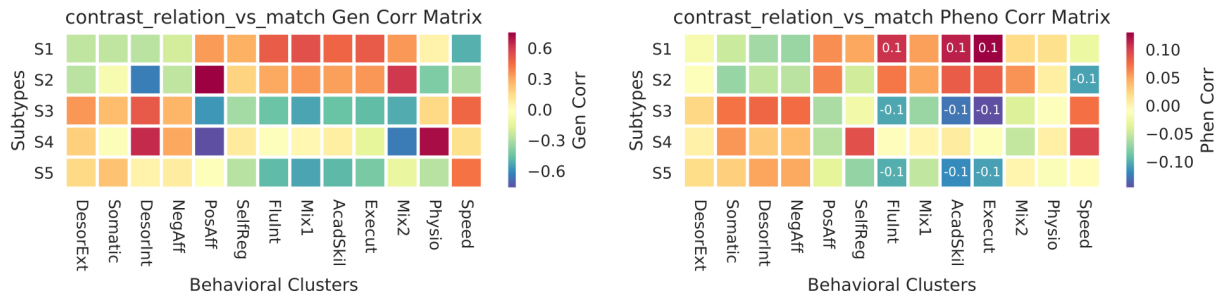


Fig. Sup 20. Genetic and phenotypic correlation for in-scanner behavioural performance: Left column represents genetic correlation between in scanner behavioural performance and brain subtypes weights. Right column represents phenotypic correlation between in scanner behavioural performance and brain subtypes weights.

Third Article.

Two highly pleiotropic mirror subtypes of brain activation shared across tasks in the Human Connectome Project

by

Yassine Benhajali[1], Daniel Pérusse[2], Moreau Clara[3], AmanPreet Badhwar[4], Ferre Perrine[5], Urchs Sebastian[6], Chouinard-Decorte Francois[7], Vainik Uku[8], and Pierre Bellec[9]

([1]) Département d'anthropologie, Université de Montréal, Montréal, Québec, Canada.

([2]) Département d'anthropologie, Université de Montréal, Montréal, Québec, Canada.

([3]) CHU Sainte-Justine Research Center, Montréal, Québec, Canada.

([4]) Centre de recherche de l'Institut universitaire de gériatrie de Montréal (CRIUGM), Montréal, Québec, Canada.

([5]) Centre de recherche de l'Institut universitaire de gériatrie de Montréal (CRIUGM), Montréal, Québec, Canada.

([6]) Montreal Neurological Institute, Montréal, Québec, Canada

([7]) Montreal Neurological Institute, Montréal, Québec, Canada.

([8]) Institute of Psychology, University of Tartu, Estonia

([9]) Département de psychologie, Université de Montréal, Montréal, Québec, Canada.

This article was submitted in Neuroimage.

Abstract.

Brain activation maps have been shown to exhibit high heritability estimates, using functional magnetic resonance. These works examined a range of different tasks independently. However little is known about common genetic influence on brain activation across different cognitive tasks.

We used the Human Connectome Project (HCP) fMRI task dataset, featuring 7 cognitive domains and 842 participants. We first grouped brain maps into subtypes of activation maps that share the same brain activation patterns. Then we estimated the genetic correlation between brain subtypes across tasks, as well as their association with behavioural measures.

Our results showed a high genetic correlation between brain subtypes from different cognitive domains as well as modular organizations of these shared genetic influences. Two main modules (meta-subtype) emerged that associate with behavioural measures, more genetically and phenotypically than environmentally. Finally, each module is linked either to a positive or negative dimension of behavioural measures.

These two dominant gradients of genetic influence on brain activation seem to be linked either to a positive or negative dimension of behavioural measures. Future work could link these gradients to gene expression.
Keywords: Task-fMRI, Subtypes, Heritability, Genetic Correlation, Behavioural measures.

1. Introduction

Brain function is highly variable across individuals, and driven in large part by genetic factors (Polderman et al., 2015). Modes of this variation ("subtypes") can be identified across individuals using cluster analysis (Drysdale et al., 2016). In previous work, we extracted brain subtypes for each of the 7 Human Connectome Project (HCP) fMRI tasks (Benhajali et al., 2020b). We investigated each task separately and we found that subtype membership for a given task has a strong genetic basis and is linked to behavioural measures inside and outside the scanner. In the present work, we question whether there is link across subtypes from different tasks and domains. We specifically aimed to explore subtypes between the 7 HCP tasks by (1) testing whether a genetic link between subtypes from unrelated cognitive domains (meta-subtypes) exists, and to (2) measure the association between meta-subtypes and behavioural measures.

1.1. Inter-individual differences in brain activation and brain activation subtypes

Task activation maps exhibit substantial inter-individual differences (Hawco et al., 2020). These differences likely reflect both differences in the anatomo-functional organization of brain networks (Llera et al., 2019; Kanai and Rees, 2011), and differences in cognitive strategies implemented by different subjects (Seghier and Price, 2018). Differences in brain activation related to either cognitive strategies or individual abilities have been demonstrated across a wide range of cognitive domains such as memory tasks (Heun et al., 2000; Feredoes and Postle, 2007; Miller et al., 2009), language task (Seghier et al., 2008) or visual task (Machielsen et al., 2000). In a study using sentence-picture verification paradigm in fMRI, Miller and colleagues (Miller et al., 2012) manipulated the use of either verbal or visual strategies in a memory retrieval task in healthy subjects. They found differences between individual brain activations patterns are largely explained by their retrieval strategies. These observations show that average group activation maps do not accurately represent the neural systems engaged by the individuals that make up that group (Lebreton and Palminteri, 2016).

Cluster analysis has the potential to reveal the inter-individual heterogeneity in brain activation. A cluster analysis summarizes the heterogeneity of brain maps using a finite number of prototype maps, or subtypes (See Figure 1). Hierarchical clustering was for instance applied to aggregate subjects based on the similarity of brain activation maps (Kherif et al., 2003), and identify atypical subjects (Seghier et al., 2007). In the non-clinical domain, some studies have linked activation subtypes with normal behavioural and demographic measures. A recent study by Kashyap and colleagues (Kashyap et al., 2019) looked at resting state brain subtypes from 788 healthy subjects from HCP. They estimated the common component of each subject from the four resting states runs, and then classified the subjects based on the similarity/dissimilarity of their components. Two subgroups emerged, one with high components similarity and another with high component dissimilarity. Both subgroups associated with behavioural measures related to the use of marijuana, illicit drugs, alcohol, tobacco, and a measure of antisocial personality. In a previous work (Benhajali et al., 2020b), we showed that the language and working memory task activation subtypes, from the HCP sample, are correlated genetically and phenotypically with behavioural measures. These studies demonstrate that heterogeneity exists between participants even in healthy control populations and could be characterized as subtypes using unsupervised data mining tools. A key question left unanswered in these previous works is: do subtypes captured from different cognitive domains have a common intrinsic functional pattern (meta-subtypes)? And, if so, is there a shared genetic basis for these meta-subtypes and individual variations in behaviour?

1.2. Genetic correlation

Despite the great literature on inter-individual differences in brain activation, and the growing literature on brain subtyping, it is still unclear whether brain activation subtypes could serve as a useful intermediate phenotype between genes and behaviour. Having established the heritability of brain activation, some studies attempted to estimate whether brain functional activation maps are under the influence of similar groups of genes as behavioural measures, a concept called pleiotropy or genetic correlation between phenotypes (Paaby and Rockman, 2013). Le Guen and colleagues (Le Guen et al., 2018) used HCP data to perform bivariate genetic analyses between the fMRI activation maps and fluid intelligence, as well the performance of subjects on tasks in the working memory and language domains. They observed that several parts of the language network along the superior temporal sulcus, as well as the angular gyrus belonging to the math processing network, are significantly genetically correlated with these indicators of cognitive performance. Rao and colleagues (Rao et al., 2018) provided evidence for the genetic correlation between risk-taking behaviour and risk-related brain activation. They found a moderate genetic correlation between risk-taking behaviour and risk-related brain activation in the left insula, right striatum, and right superior parietal lobule. In previous work (Benhajali et al., 2020b), we used HCP task-fMRI data to identify subtypes of brain activation and test their phenotypic and genetic association with behavioural measures. We found that subtypes from 7 HCP tasks were moderately to highly heritable, with the language and memory task subtypes genetically correlated with behavioural measures. A key question remains largely open: do subtypes across different cognitive domains share the same genetic influences? And how do they associate with behavioural measures?

1.3. Aims and hypotheses

In this work, by combining subtypes from different cognitive domains (7 HCP tasks) and exploring their genetic association, our specific aims and hypothesis are:

(1) Estimate the genetic correlation between subtypes from different cognitive domains and test the existence of modular genetic influence on brain subtypes across the 7 HCP tasks by clustering techniques. Our hypothesis was that subtypes from different cognitive domains will show high genetic correlation, and will cluster into modular meta-subtypes.

(2) Estimate the genetic, environmental and phenotypic correlation between weights meta-subtypes and behavioural measures. Our hypothesis is that weight meta-subtype will show an association of genetic, environmental or phenotypic nature with behavioural measures outside of the scanner.

2. Method

2.1. Subjects

We used the S900 data release from the HCP open access database (Van Essen et al., 2013). The S900 release includes behavioural and imaging data from 897 healthy adult participants collected between August 2012 and March 2015, out of which 862 healthy adults (aged 22–35 y) successfully completed at least one of the 7 fMRI tasks. Following quality control of the preprocessed imaging data, we selected 842 healthy adults (composed of 742 individuals from 278 families, and 100 unrelated individuals) for inclusion in the analysis. Individuals from the same family were either monozygotic twins (genetically identical, 84 pairs, 168 individuals) or dizygotic/fraternal twins (genetically no more related than ordinary full siblings, 77 pairs, 154 individuals) and 89 families of non-twin siblings (209 individuals) (Van Essen et al., 2013).

The HCP sample reflects the ethnic diversity of American families (White non-Hispanic, Hispanic, Asian and African-American). HCP provides the required ethics and consent needed for study and dissemination. This study went through further internal institutional data analysis approval.

2.2. Imaging data

MRI data for the S900 release was collected on a customized Siemens MAGNETOM Connectom 3T scanner at Washington University. The data are composed of multiple imaging sessions covering four modalities: structural (T1w and T2w), resting-state fMRI, task fMRI (fMRI, 7 tasks) and diffusion MRI (dMRI, (Hodge et al., 2016). In this work we only used T1w and fMRI modalities. For T1w images a 32 min was spent acquiring each MPRAGE image. An MPRAGE sequence with 0.7 mm isotropic resolution (FOV = 224 mm, matrix = 320, 256 sagittal slices in a single slab) (Mugler and Brookeman, 1990), TR = 2400 ms, TE = 2.14 ms, TI = 1000 ms, FA = 8°, Bandwidth (BW) = 210 Hz per pixel, Echo Spacing (ES) = 7.6 ms (see (Glasser et al., 2013) for more detail). For each subject and task, 2 runs of fMRI time series data were collected, 1 run in each phase encoding direction, left-to-right (LR) then right-to-left (RL), with a TR of 0.73 s and a spatial resolution of 2-mm isotropic, using a standard 32-channel Siemens coil (Van Essen et al., 2013)

All included participants completed 7 tasks in the MRI scanner - Working Memory (405 frames/scan), Gambling (253 frames/scan), Motor (284 frames/scan), Language (316 frames/scan), Social Cognition (274 frames/scan), Relational Processing (232 frames/scan),

and Emotion Processing (176 frames/scan). An in depth description of the experimental design can be found in the original publication (Hodge et al., 2016).

2.3. Image preprocessing

From HCP S900 data release, we used the unprocessed tfMRI data. Each subject was preprocessed using the NeuroImaging Analysis Kit version dev-0.14.0 (NIAK[10]), under CentOS with Octave[11] version 3.8.1 and the Minc toolkit[12] version 0.3.18. Analyses were executed in a high-performance computing environment[13], using the pipeline system for Octave and Matlab (PSOM) (Bellec et al., 2012).

Each tfMRI image was corrected for differences in timing of slice acquisitions; a rigid-body motion was then estimated using Minctracc (Collins and Evans, 1997) for each time frame, both within and between runs, as well as between one fMRI run and the T1 scan for each subject. The T1 scan was itself non-linearly co-registered to the Montreal Neurological Institute (MNI) ICBM152 stereotaxic symmetric template (Fonov et al., 2011a), using the CIVET pipeline (Ad-Dab'bagh et al., 2006). The rigid body, fMRI-to-T1 and T1-to-stereotaxic transformations were all combined to resample the fMRI in MNI space at a 3 mm isotropic resolution. To minimize artifacts due to excessive motion, all time frame showing an average frame displacement (FD) greater than 0.5 mm were removed (Power et al., 2012). The following nuisance covariates were regressed out from the tfMRI time series: slow time drifts (basis of discrete cosines with a 0.01 Hz high-pass cut-off), average signals in conservative masks of the white matter and the lateral ventricles as well as the first principal components (accounting for 95% variance) of the six rigid-body motion parameters and their squares (Giove et al., 2009). Finally, the tfMRI volumes were spatially smoothed with a 6 mm isotropic Gaussian blurring kernel. A more detailed description of the pipeline can be found on the NIAK website[14].

2.4. Registration and motion quality control

We used a structured visual quality control procedure (Benhajali et al., 2020a) to assess the registration between the individual T1 scans and the stereotaxic template space, as well as registration between the individual T1 and the median functional image. The quality control procedure included a series of key anatomical landmarks and associated confidence intervals. The quality control procedure is publicly available as a tutorial on the

[10]http://niak.simexp-lab.org/
[11]https://www.gnu.org/software/octave/
[12]http://bic-mni.github.io/
[13]https://www.calculquebec.ca/
[14]http://niak.simexp-lab.org/build/html/PREPROCESSING.html

Zooniverse platform (noa, 2015a). Subjects were excluded from the analysis in cases of failed coregistration with the template space or were fewer than 100 time points had acceptable amounts of head motion after scrubbing (see next section). A total of 843 subjects (477 female) passed quality control criteria for inclusion in the analysis.

In scanners head motion is a critical confounding factor in functional MRI and tends to act as a relatively stable, heritable phenotype (Couvy-Duchesne et al., 2014; Engelhardt et al., 2017; Couvy-Duchesne et al., 2016). Removing time points affected by large head motion, a technique known as scrubbing, is efficient at reducing the effect of head motion on fMRI time series (Caballero-Gaudes and Reynolds, 2017). Average framewise displacement (FD) (Power et al., 2012) is a summary measure of head displacement throughout an fMRI acquisition, and can be estimated both before and after scrubbing. We assessed the heritability of average FD motion measure before and after removal of time frames with excessive motion (time frames with FD greater than 0.5 mm). Heritability analysis was implemented using the SOLAR Eclipse software, as described in the dedicated Methods section below. Figure Sup 24 (supplementary material) shows that FD measures before scrubbing had substantial (and significant) heritability estimates, ranging from 0.22 to 0.52 across tasks, which is consistent with previous reports (Couvy-Duchesne et al., 2014; Engelhardt et al., 2017; Couvy-Duchesne et al., 2016). After scrubbing, the heritability estimates of FD ranged from 0 to 0.13 across tasks. We concluded that using scrubbing as a preprocessing step was beneficial for further heritability analysis of fMRI activation maps, as it reduced the potential confounding effects of heritable motion artifacts.

2.5. Activation maps

For each of the 7 tasks, individual activation maps were extracted from block design conditions using the statistical analysis toolbox FMRISTAT (Worsley et al., 2003). Different conditions were included in some of the tasks (e.g. foot, hand and tongue for the motor task), resulting in a total of 21 conditions. For each of the 7 tasks, the stimuli in the design matrices were convolved with a hemodynamic response function modelled as a difference of two gamma functions. Temporal drift was removed by adding a cubic spline in the frame times to the design matrix, and spatial drift was removed by adding a covariate in the whole volume average. The autocorrelation parameters were estimated at each voxel and used to whiten the data and design matrix.

2.6. Meta-subtypes from the 7 HCP tasks

For each of the 21 tasks and conditions, a nuisance regression is applied on stacked activation maps (Figure 21a). Nuisance regressors are frame displacement (FD), systolic

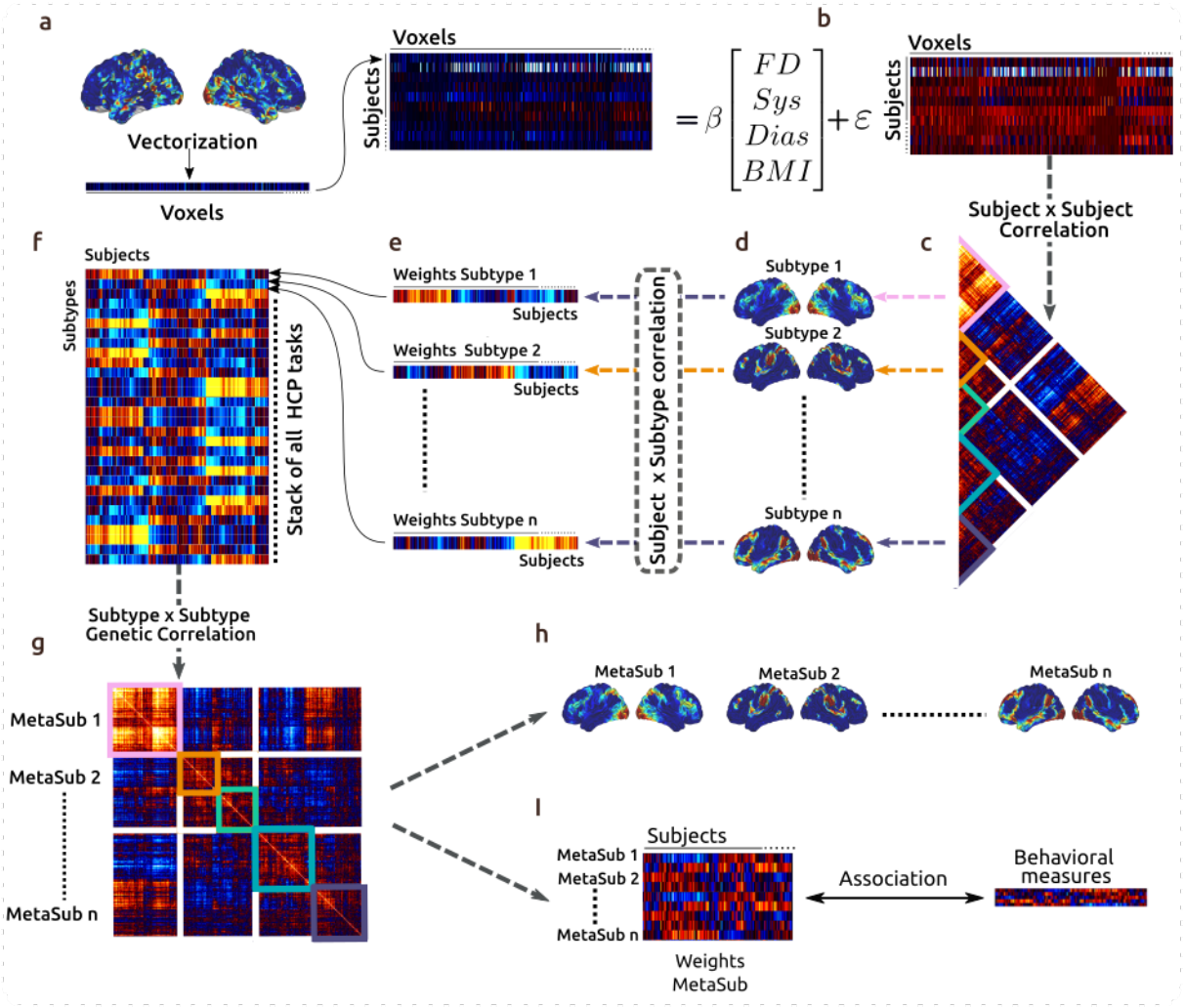


Fig. 21. meta-subtypes generation and association with behavioural measures- **a**From each subject, an un-thresholded statistical parametric map (SPM) is vectorized then stacked. **b**. Nuisance regression from stacked SPM (FD: Frame Displacement, Sys: Systolic blood pressure, Dias: Diastolic blood pressure, BMI: Body Mass Index). **c**. From the residual stack, a subject by subject correlation matrix is clustered into subgroups of activations maps (subtypes). **d**. Each subtype is represented spatially by a mean activation map of all subjects within the same cluster . **e**. Each subtype is represented on the subject level by its weight. The subtypes weight is the correlation between a vectorized subject activation map and a vectorized subtype activation map. **f**. All subtypes weights from the 7 tasks are stacked into subjects by subtypes' weight matrix. **g**. A subtypes weight by subtypes weight genetic correlation matrix is clustered into subgroups of genetic meta-subtypes. **h**. Each meta-subtype is represented spatially by a mean map of all subtypes within the same cluster. **i**. Mean weights of genetic meta-subtypes are genetically, environmentally and phenotypically associated with behavioural measure.

blood pressure, diastolic blood pressure and body mass index (Figure 21b). Then, a subject by subject similarity (Pearson's correlation) matrix summarized the between-subject similarity of activation maps (Figure 21c). Next, a hierarchical cluster analysis was applied

on all individual activation maps, which identified homogeneous subgroups of maps for each condition. Within each activation map subgroup, we estimated the average activation map across all subgroup members, or subtype map (Figure 21d). We repeated steps (a) to (d) for each task and condition in order to generate 5 subtype maps from each task and condition for a total of 105 subtypes brain maps.

We compute the spatial similarity (correlation) of each individual activation map in the discovery sample with each of the identified activation subtypes (Figure 21e). The estimated spatial correlation coefficient is called the subtype weight and each individual has one for each subtype, ranging from -1 (perfect anticorrelation of the individual and the subtype map) to +1 (perfect correlation of the individual and subtype map). The subtype weights represent a continuous measure of similarity between a given individual activation map and a given subtype map. Unlike discrete assignments made by the hierarchical clustering, the weights are a soft assignment and can be seen as a dimension reduction of the individual maps, rather than pure categorical summary. It has been argued by several prior studies that such continuous summaries are more appropriate than discrete assignments (Zhang et al., 2016; Badhwar et al., 2019b).

2.7. Meta-subtype generation and association with behavioural measures

After generating individual subtypes weights for each task separately (Figure 21e), we stack all subtypes weights across the 7 types of cognitive tasks into subjects by subtypes' weight matrix (Figure 21f). Then, we estimate genetic correlation between all subtypes and cluster the resulting correlation into subgroups of genetic meta-subtypes (Figure 21g). Each meta-subtype is represented spatially by a mean map of all subtypes within the same cluster (Figure 21h). Finally, the mean weights of genetic meta-subtypes are genetically, environmentally and phenotypically associated with behavioural measures (Figure 21h-j). We estimated genetic, phenotypic and environmental correlation between meta-subtypes weights and behavioural measure (Figure 21h-j) using the Sequential Oligogenic Linkage Analysis Routines Eclipse (SOLAR) software package (Almasy and Blangero, 1998). We used equation (2.1) to simultaneously quantify the shared genetic variance and the phenotypic correlation between brain phenotypes (weights) and behavioural measures (in and out-of-scanner behavioural measures). To assess this relationship, we used SOLAR, relying on the following equation:

$$\rho_p = \sqrt{h_a^2}\sqrt{h_b^2} \cdot \rho_g + \sqrt{1 - h_a^2}\sqrt{1 - h_b^2} \cdot \rho_e \tag{2.1}$$

Where Pearson's phenotypic correlation ρ_{p} is decomposed into $\rho_{\mathrm{g}} and \rho_{\mathrm{e}}$. ρ_{g} is the proportion of variability due to shared genetic effects and ρ_{e} that due to the environment, while h_a^2 and h_b^2 correspond to the previously defined narrow sense heritability for phenotypes a and b, respectively. In our case, one corresponds to the heritability of subtypes weight while the second is the heritability of one of our behavioural scores.

3. Results

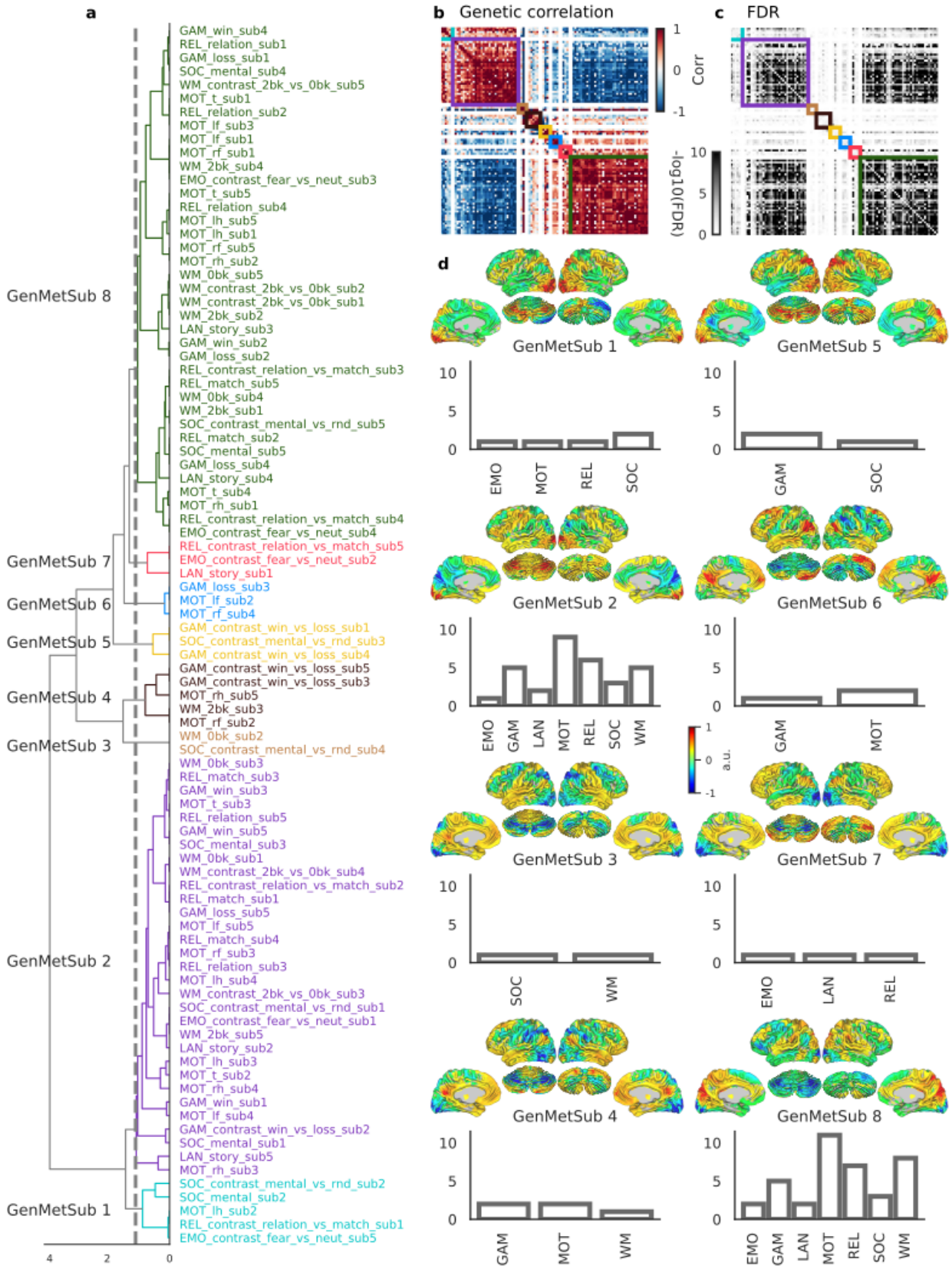

a
GAM_win_sub4
REL_relation_sub1
GAM_loss_sub1
SOC_mental_sub4
WM_contrast_2bk_vs_0bk_sub5
MOT_t_sub1
REL_relation_sub2
MOT_lf_sub3
MOT_lf_sub1
MOT_rf_sub1
WM_2bk_sub4
EMO_contrast_fear_vs_neut_sub3
MOT_t_sub5
REL_relation_sub4
MOT_lh_sub5
MOT_lh_sub1
MOT_rf_sub5
MOT_rh_sub2
WM_0bk_sub5
WM_contrast_2bk_vs_0bk_sub2
WM_contrast_2bk_vs_0bk_sub1
WM_2bk_sub2
LAN_story_sub3
GAM_win_sub2
GAM_loss_sub2
REL_contrast_relation_vs_match_sub3
REL_match_sub5
WM_0bk_sub4
WM_2bk_sub1
SOC_contrast_mental_vs_rnd_sub5
REL_match_sub2
SOC_mental_sub5
GAM_loss_sub4
LAN_story_sub4
MOT_t_sub4
MOT_rh_sub1
REL_contrast_relation_vs_match_sub4
EMO_contrast_fear_vs_neut_sub4
REL_contrast_relation_vs_match_sub5
EMO_contrast_fear_vs_neut_sub2
LAN_story_sub1
GAM_loss_sub3
MOT_lf_sub2
MOT_rf_sub4
GAM_contrast_win_vs_loss_sub1
SOC_contrast_mental_vs_rnd_sub3
GAM_contrast_win_vs_loss_sub4
GAM_contrast_win_vs_loss_sub5
GAM_contrast_win_vs_loss_sub3
MOT_rh_sub5
WM_2bk_sub3
MOT_rf_sub2
WM_0bk_sub2
SOC_contrast_mental_vs_rnd_sub4
WM_0bk_sub3
REL_match_sub3
GAM_win_sub3
MOT_t_sub3
REL_relation_sub5
GAM_win_sub5
SOC_mental_sub3
WM_0bk_sub1
WM_contrast_2bk_vs_0bk_sub4
REL_contrast_relation_vs_match_sub2
REL_match_sub1
GAM_loss_sub5
MOT_lf_sub5
REL_match_sub4
MOT_rf_sub3
REL_relation_sub3
MOT_lh_sub4
WM_contrast_2bk_vs_0bk_sub3
SOC_contrast_mental_vs_rnd_sub1
EMO_contrast_fear_vs_neut_sub1
WM_2bk_sub5
LAN_story_sub2
MOT_lh_sub3
MOT_t_sub2
MOT_rh_sub4
GAM_win_sub1
MOT_lf_sub4
GAM_contrast_win_vs_loss_sub2
SOC_mental_sub1
LAN_story_sub5
MOT_rh_sub3
SOC_contrast_mental_vs_rnd_sub2
SOC_mental_sub2
MOT_lh_sub2
REL_contrast_relation_vs_match_sub1
EMO_contrast_fear_vs_neut_sub5
GenMetSub 8
GenMetSub 7
GenMetSub 6
GenMetSub 5
GenMetSub 4
GenMetSub 3
GenMetSub 2
GenMetSub 1
4
2
0
b Genetic correlation
c FDR
1
0
-1
Corr
10
5
0
-log10(FDR)
d
GenMetSub 1
GenMetSub 5
GenMetSub 2
GenMetSub 6
GenMetSub 3
GenMetSub 7
GenMetSub 4
GenMetSub 8
a.u.
EMO
MOT
REL
SOC
GAM
LAN
WM

Fig. 22. Genetic meta-subtypes - a - Dendrogram of clustered subtypes: Subtypes from different tasks are clustered into 8 meta subtypes according to the distance of their genetic correlation. **b - Genetic correlation matrix**: The matrix of genetic correlation between weights subtypes from all 7 tasks shows 8 clusters of genetic meta-subtypes (GenMetSub1-8). **c - FDR estimate for genetic correlation**: FDR estimate shows that GenMetSub 2 and 8 are the most significant clusters. **d - Spatial distribution of genetic Meta-Subtypes**: Spatial meta-subtypes maps composed of the man subtypes maps that constitute each group.

3.1. Subtypes are genetically correlated across tasks and cluster into meta-subtypes

The main goal of this work was to assess the genetic correlation between subtypes across the 7 HCP tasks. First, we estimated the genetic correlation between each pair of subtypes' weights from all 7 tasks. Then, we used hierarchical clustering technique on the matrix of genetic correlation to reveal the underlying organization. The resulting matrix of genetic correlation (Figure 22b) was strikingly organized into two clusters (GenMetSub 2 and 8), with high values of negative and positive associations (-1 to 1) between the two clusters. These two genetic meta-subtypes were the most significant (figure 22c) and both composed of a mixture of subtypes weights coming from all the 7 tasks (Figure 22a).

These meta-subtypes can be visualized on the spatial domain by averaging the subtypes maps that make up each meta-subtype. Figure 22d shows the spatial maps of each meta-subtype and their composition in terms of different tasks wise composition. GenMetSub 2 and 8 shows the most heterogeneous task composition and are spatially mirrored versions of each other.

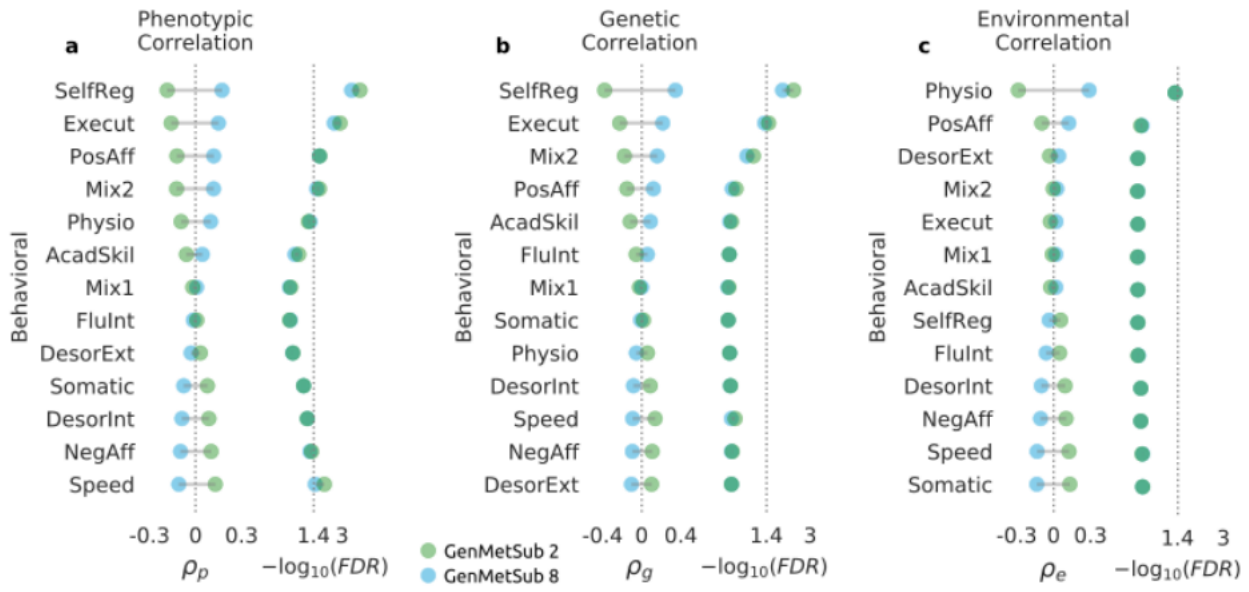


Fig. 23. Genetic meta-subtypes association with behavioural measures - a - Phenotypic association between genetic meta-subtypes and behavioural measures: Phenotypic correlation between genetic meta-subtypes (2,8) and clusters of behavioural measures. GenMetSub 2 and 8 shows a significant association with behavioural clusters related to self-regulation, executive function, positive effects and speed processing. **b - Genetic association between genetic meta-subtypes and behavioural measures:** Genetic correlation between genetic meta-subtypes (2,8) and clusters of behavioural measures. GenMetSub 2 and 8 shows a significant association with behavioural clusters related to self-regulation and executive functions. **c - Environmental association between genetic meta-subtypes and behavioural measures:** Environmental correlation between genetic meta-subtypes (2,8) and clusters of behavioural measures. GenMetSub 2 and 8 shows a no significant association with behavioural clusters.

3.2. Genetic meta-subtypes are associated with Behavioural measures

The second goal of this work is to expose the genetic and environmental link between meta-subtypes and behavioural measures. We estimated the genetic, phenotypic and environmental correlation between GenMetSub 2 and behavioural measures, then we did the same with GenMetSub 8. Figure 23a shows phenotypic correlation between genetic meta-subtypes (GenMetSub 2, 8) and clusters of behavioural measures. SelgReg have the highest and most significant association with both GenMetSub 2 and 8 ($\rho_p = 0.25$), followed by Execut, PosAff and Mix2 ($\rho_p = 0.23, \rho_p = 0.20 and \rho_p = 0.19$). Since phenotypic correlation is known to follow the genetic correlation (Cheverud, 1988; Sodini et al., 2018; van Rheenen et al., 2019), our results in figure 23b show the same trend were genetic correlation between genetic meta-subtypes (GenMetSub 2,8) and behavioural measures is also high for Execut

($\rho_g = 0.37$) and SelfReg ($\rho_g = 0.25$). Finally, the environmental correlation between genetic meta-subtypes and behavioural measure are nonsignificant except for the cluster Physio ($\rho_e = 0.32$).

4. Discussion

Inter-individual variability in brain activation is partly explained by genetic and environmental variability (Elliott et al., 2018). However, it is unclear how this variability is organized across cognitive domains. In this work, three main findings: first, we showed genetic correlation between subtypes from seven cognitive domains (emotional, language, relational, social, memory, motor and risk taking). Second, these correlations between subtypes were clustered into two main meta-subtypes that were anticorrelated, showing mirror activation patterns. Finally, these core meta-subtypes were associated with behavioural measures.

4.1. Genetic correlation between subtypes

Brain subtypes (from 7 HCP tasks) used as brain endophenotypes are an effective way to capture variability in the brain (Benhajali et al., 2020b). These subtypes appear to have a biological grounding (high heritability). We extended these results by showing that these cognitive domains brain subtypes share a common genetic component, suggesting pleiotropy mechanisms.

Most studies used pleiotropy to link brain functional phenotype (activation or resting state) to behavioural measures (Le Guen et al., 2018). In our work, we estimated pleiotropy between subtypes from different domains. We showed that pleiotropy is high within subtypes from the same task as well as between tasks. This finding is consistent with gene expression studies on brain networks (Richiardi et al., 2015; Patania et al., 2019). These studies rely on the mapping of microarray-based gene expression data from human brain samples onto functional and anatomical brain networks. A study by Bertolero and colleagues (Bertolero et al., 2019) combined gene expression from six postmortem brains from the Allen Brain Institute, functional and structural brain connectivity, genotyping, and cognitive and behavioural testing of more than 1000 young adults. They identified a modular pleiotropic genetic map of the human brain's network architecture. A modular pleiotropy occurs when the genotypes-phenotypes relationship typically is determined by mostly non-overlapping sets of genes (Wagner and Zhang, 2011). This modular pleiotropy was also reported to be more strongly related to functional connectivity than to structural connectivity. Our results in figure 22b might also relate to this concept.

4.2. Across tasks meta-subtypes

After estimating genetic correlation between all subtypes of the 7 HCP tasks, we clustered them according to their Euclidean distance. This technique captures aggregation patterns across all 7 tasks. We identified eight meta-subtypes that aggregate heterogeneous subtypes from the 7 HCP tasks. Two significant meta-subtypes emerged: GenMetSub2 and GenMetSub8. Each of these two clusters contains at least one subtype from each of the 7 hcp tasks (figure 22a). This finding is consistent with the modularity concept described in the section above. The heterogeneity of the two modules are, furthermore, inconsistent with the pleiotropic modularity observed in brain connectivity studies (Richiardi et al., 2015; Bertolero et al., 2015, 2018; Patania et al., 2019). The modularity observed in Bertolero studies (Bertolero et al., 2018, 2019) showed that gene expression is organized into mostly non-overlapping sets, where each set encodes the connectivity of a single brain network community. Specifically, the coexpression of a particular gene set could statistically explain the connectivity of one brain community but not the connectivity of any other brain community. The modularity observed in our study shows heterogeneous task composition (Figure 22c) on the two significant modules (GenMetSub 2 and 8). Modularity might be state-specific (resting state based networks versus task-based networks). Brain networks exhibit community structure that reconfigures during cognitively demanding tasks (Cole et al., 2014). Betzel and colleagues (Betzel et al., 2018) studied the community structure of functional brain networks while subjects either rest or perform tasks. They found evidence that the resting brain is largely assortative, while this assortative structure breaks down during tasks and is supplanted by core, periphery, and disassortative communities. This may help to disentangle the cognitively heterogeneous composition of our 2 modules.

4.3. Association with Behavioural measures

Our final goal was to link the meta-subtypes to behavioural measures. We estimated the phenotypic, genetic and environmental correlation between the significant pleiotropic modules (GenMetSub 2, 8) and behavioural measures. The results show mirror associations between the two pleiotropic modules. When sorting behavioural association with GentMetSub 2 and 8 from the highest to the lowest, a positive-negative axis appeared with as nearly all the positively correlated behaviours are commonly considered as positive personal qualities or indicators (for example, positive affect, executive performance, academic achievement), and all negatively correlated behaviour relate to negative traits (for example, internalized and externalized disorder, negative affect, somatic). This finding is consistent with the work of Smith and colleagues (Smith et al., 2015), where they found one significant mode (positive-negative) of population variation that links a specific pattern of brain connectivity to a specific pattern of covariance between many behavioural and demographic subject

measures. Importantly, the work of Smith and colleagues had only considered resting-state fMRI, and did not consider task activation maps. They also relied on a different type of modelling (factor analysis). Like in our result for the GenMetSub 8, the vast majority of the subject's measures that correlate positively with this mode where positive subject traits and measures; those that correlate negatively are mostly negative subject measures. For GenMetSub 2, we observed the same relation with behaviour but the association is a mirrored correlation value compared to GenMetSub 8. These results suggest that the modular pleiotropy observed in brain activation maps is associated with two modes of population on the behavioural level (positive-negative).

4.4. Limitations

This study has some limitations. First, in this work we made the choice of preprocessing the HCP data and not using the HCP preprocessed data mainly to implement motion censoring. Since head motion is highly heritable, we opted for using the scrubbing method to ensure that time points with excessive motion are removed (Figure Sup1). Despite our efforts to reduce the effects of head motion, we still need to be careful in the interpretation of our result, since some effects of head motion may remain in the data and bias heritability estimate (Bolton et al., 2020).

Second, in this work, we could not accurately estimate the effect of the shared environment between siblings, since HCP did not provide the household information necessary to model it. We opted for using the mother ID as a household ID. This is not ideal, since split families (ex: divorced) are still considered to share the same familial environments, which is not necessarily the case. Our heritability estimates could be slightly overestimated for that reason.

Third, we should caution about the possibly limited generalizability of our study since the majority of participants were Caucasian (European origin). HCP did offer information about the ethnicity of the participants but this information was not sufficient to use as confounding elements in our analysis.

4.5. Future directions

Our work on healthy populations could explain better the nature of those subgroups in the normal population by pointing to the biological underpinning of their manifestation, and their relationship with behaviour. As a future direction, one could compare the distribution of the meta-subtypes of activation maps with gene expression in the Allen brain atlas (Sunkin et al., 2013).

5. Conclusions

Two major subtypes of brain activation with high pleiotropic association can be found across cognitive domains. These meta-subtypes of activation also show high levels and mirror pleiotropic associations with behaviour. Individual maps of task activation follow simple principles of variations, both spatially and genetically. This suggests that a large portion of inter-individual variations in brain activity and behaviour could be explained by a common set of genes, irrespective of the type of cognitive tasks under consideration.

Supplementary material

Task fMRI

Working Memory. N-back tasks were used as follows: the participant was presented with a series of pictures, drawn from different categories (faces, tools, places and body parts). Half of the blocks were used for a 2-back memory trial, where the participant was asked to respond each time whether the current stimulus was the same as the one from two back. The other half of the blocks were used for a 0-back working memory task, where the individual had to recognize a single 'target' for any stimulus during the block.

Incentive Processing (gambling). Participants played a game of card guessing, and were asked to guess the number on a mystery card (question mark card) to win or lose money. The potential number on the mystery card was an integer between 1 and 9. Pressing one of two buttons, the participants had to guess if the number was over or below 5. A successful guess was rewarded by 1$, the participants lost 0.50$ for bad guesses, and nothing for neutral guesses (where the number was exactly 5). The task was presented in blocks of 8 trials that were either mostly reward (6 reward trials pseudo randomly interleaved with either 1 neutral and 1 loss trial, 2 neutral trials, or 2 loss trials) or mostly loss (6 loss trials interleaved with either 1 neutral and 1 reward trial, 2 neutral trials, or 2 reward trials). In each of the two runs, there were 2 mostly reward and 2 mostly loss blocks, interleaved with 4 fixation blocks (15 s each)

Motor. Participants were presented with visual cues that asked them to tap their left or right fingers, squeeze their left or right toes, or move their tongue, in order to generate maps of different motor areas. Each block of a movement type lasted 12 s (10 movements), and was preceded by a 3 seconds cue. In each of the two runs, there were 13 blocks, with 2 blocks of each body part and three blocks of 15 s fixation.

Language Processing. The task consisted of two runs that each interleave 4 blocks of a story condition and 4 blocks of a math condition. The task was designed so that the math blocks match the length of the story blocks, without fixation between block transitions. The story blocks presented participants with brief auditory stories (5–9 sentences) adapted

from Aesop's fables, followed by asking participants to select the topic of the story among two alternatives (button press response). The math blocks required subjects to complete addition and subtraction problems (auditory instructions, and select the right answer among two alternatives (button press response).

Social Cognition (Theory of Mind). Participants were presented with short video clips (20 s) of objects (squares, circles, triangles) either interacting in some way, or moving randomly. Participants were asked to judge whether a social or a random interaction was happening between these objects on the video. Each of the random and social conditions had 5 video blocks (2 Mental and 3 Random in one run, 3 Mental and 2 Random in the other run) and 5 fixation blocks (15 s each).

Relational Processing. In the relational processing condition, participants were presented with 2 pairs of objects, with one pair at the top of the screen and the other pair at the bottom of the screen. They were told that they should first decide what dimension differed across the top pair of objects (shape or texture) and then they should decide whether the bottom pair of objects also differed along that same dimension.

Emotion Processing. Participants were presented with blocks of trials that either asked them to decide which of two faces presented on the bottom of the screen matched the face at the top of the screen, or which of two shapes presented at the bottom of the screen matched the shape at the top of the screen. The faces had either angry or fearful expressions.

For more details on tfMRI in the HCP, see Barch and colleagues (Barch et al., 2013).

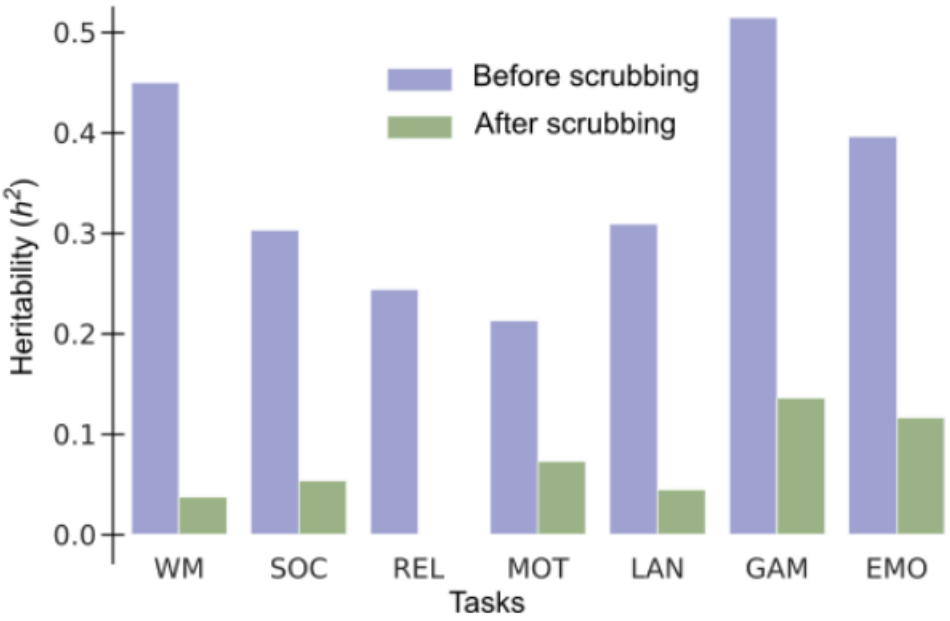


Fig. Sup 24. Heritability of framewise displacement (FD) measure before and after scrubbing on the 7 tasks. WM: working memory, SOC: social, REL: relational, MOT: motor, LAN: language, GAM: gambling and EMO: emotional. Heritability of FD was estimated before and after scrubbing to shows how motion reduction used in this paper is efficient in reducing the heritability estimate that is known to affect further estimate of heritability on brain phenotypes.

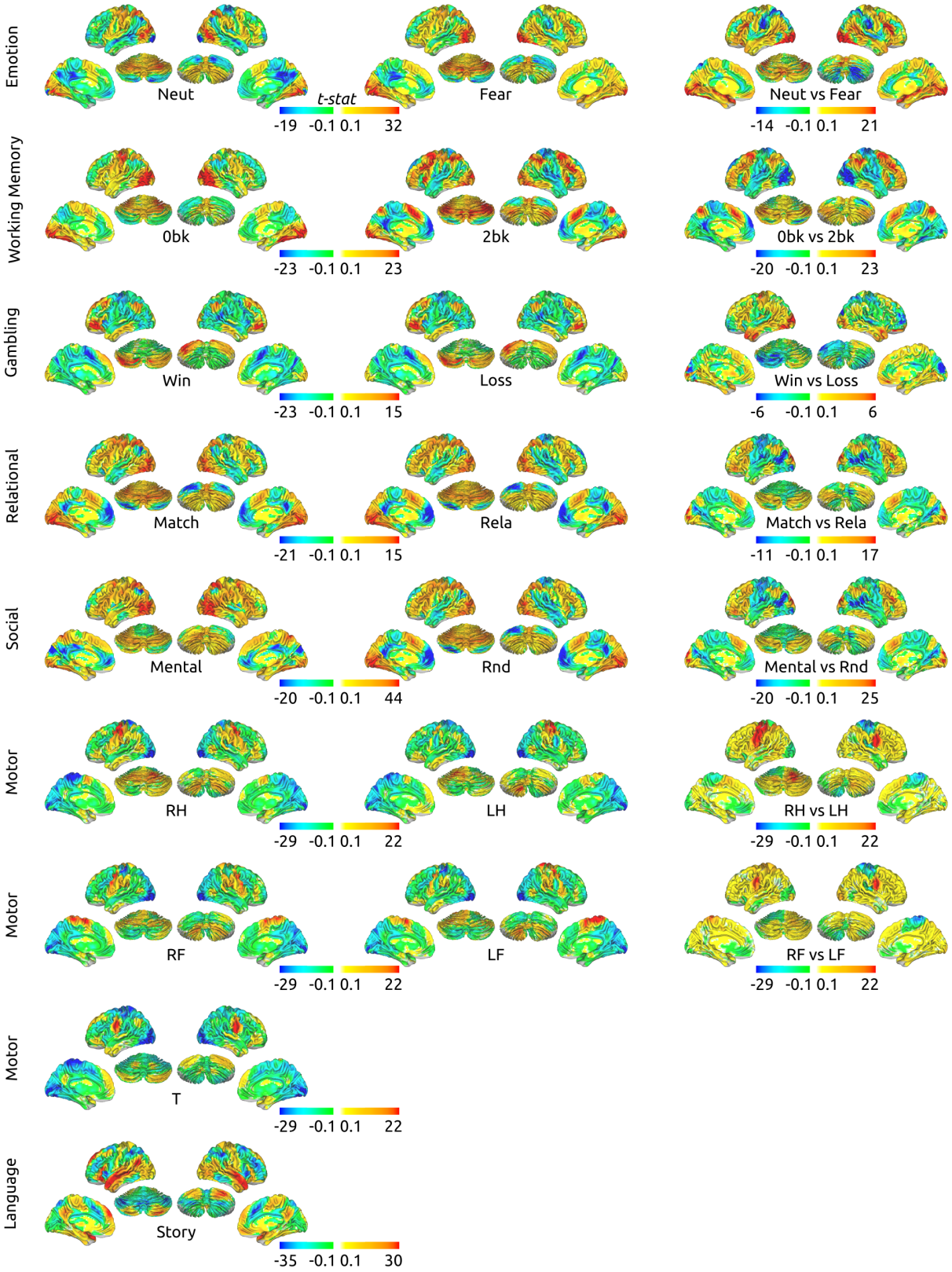

Emotion
Neut
t-stat
-19 -0.1 0.1 32
Fear
Neut vs Fear
-14 -0.1 0.1 21
Working Memory
0bk
-23 -0.1 0.1 23
2bk
0bk vs 2bk
-20 -0.1 0.1 23
Gambling
Win
-23 -0.1 0.1 15
Loss
Win vs Loss
-6 -0.1 0.1 6
Relational
Match
-21 -0.1 0.1 15
Rela
Match vs Rela
-11 -0.1 0.1 17
Social
Mental
-20 -0.1 0.1 44
Rnd
Mental vs Rnd
-20 -0.1 0.1 25
Motor
RH
-29 -0.1 0.1 22
LH
RH vs LH
-29 -0.1 0.1 22
Motor
RF
-29 -0.1 0.1 22
LF
RF vs LF
-29 -0.1 0.1 22
Motor
T
-29 -0.1 0.1 22
Language
Story
-35 -0.1 0.1 30

Fig. Sup 25. Group activation maps for the 7 HCP tasks - Emotion task: Group activation maps for neutral condition (Neut), fear condition (Fear) and the contrast Neut vs Fear. Working memory task: Group activation maps for 0-back (0bk), 2-back (2bk) and the contrast 0bk vs 2bk. Gambling task: Group activation maps for win (Win), loss (Loss) and the contrast Win vs Loss. Relational task: Group activation maps for Match, relational (Rel) and the contrast Match vs Rel. Social task: Group activation maps for Mental, random (Rnd) and the contrast Mental vs Rnd. Motor task: Group activation maps for left hand (LH), right hand (RH), left foot (LF), right foot (RF), tongue (T), contrast RF vs LF and contrast RH vs LH. Language task: Group activation maps for story.

Spatial meta-subtypes

Spatial meta-subtypes generation

After identifying 5 subtypes maps for each of the 18 contrasts (Figure 21d), we reached a total of 90 subtypes brain maps from the 7 tasks. Then, a subtype by subtype similarity (Pearson's correlation) matrix summarized the between-subtypes similarity of activation maps. Next, a hierarchical cluster analysis was applied on all subtypes activation maps from all tasks, which identified homogeneous subgroups of subtypes maps across all tasks (Figure 21f). Then, we generated a mean brain map from each subgroup, called spatial meta-subtypes maps. We compared these metasubtypes maps with a multi-resolution parcellation of functional brain networks (MIST) (Urchs et al., 2017). We used Dice coefficient (Sørensen et al., 1948) to compare the overlap between our spatial meta-subtypes maps and the MIST parcellations.

Meta-subtypes spatially overlap with resting state networks

The first goal was to explore the spatial distribution of meta subtypes from the 7 HCP task and compare them with resting state networks. We clustered subtypes of activation maps from the 7 HCP tasks in order to form subgroups of subtypes that share similar spatial distribution between all HCP tasks and conditions. A meta-subtype map was defined as the average activation map of each subgroup (Figure 21f). Figure Sup 26 shows 7 meta-subtypes maps and their composition in terms of the tasks that make up the subgroups. The resulting distribution shows that meta-subtypes are very heterogeneous in terms of tasks composition. This may suggest that meta-subtypes grouping is not task driven. In the figure Sup 26 we tested if meta-subtypes grouping is driven by intrinsic brain activation by comparing them with the resting state network from MIST parcellation. Figure Sup 26 shows substantial spatial overlap between the meta-subtypes and the resting state parcellations. The dice

coefficients represent the similarity between the spatial meta-subtypes distributions and the MIST parcellations. We estimated the dice coefficients between MIST parcellation and both activation and deactivation meta-subtypes distribution (Figure Sup 26a,b). The highest spatial overlap between the MIST parcels and the meta-subtypes are the visual networks, the default mode network and the fronto-parietal network (Figure Sup 26c).

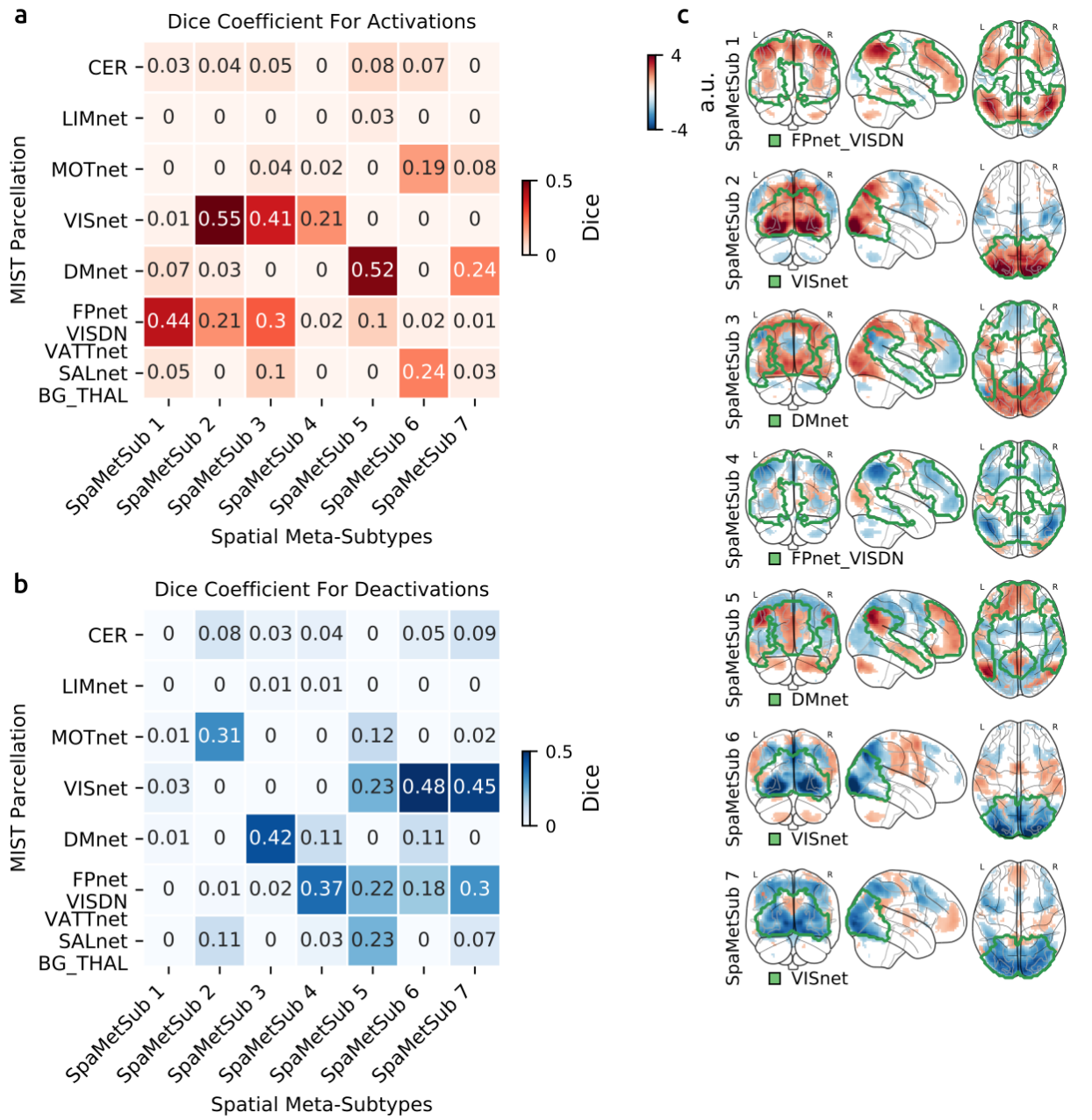


Fig. Sup 26. Spatial meta-subtypes networks - a - Dice activation overlap with MIST networks: Dice coefficient for overlap between activated brain region from meta-subtypes and the MIST parcellation. **b - Dice deactivation overlaps with MIST networks:** Dice coefficient for overlap between deactivated brain region from meta-subtypes and the MIST parcellation. **c - MIST and meta-subtypes maps overlap:** Spatial overlap between MIST networks scale 7 and the 7 spatial meta-subtypes.

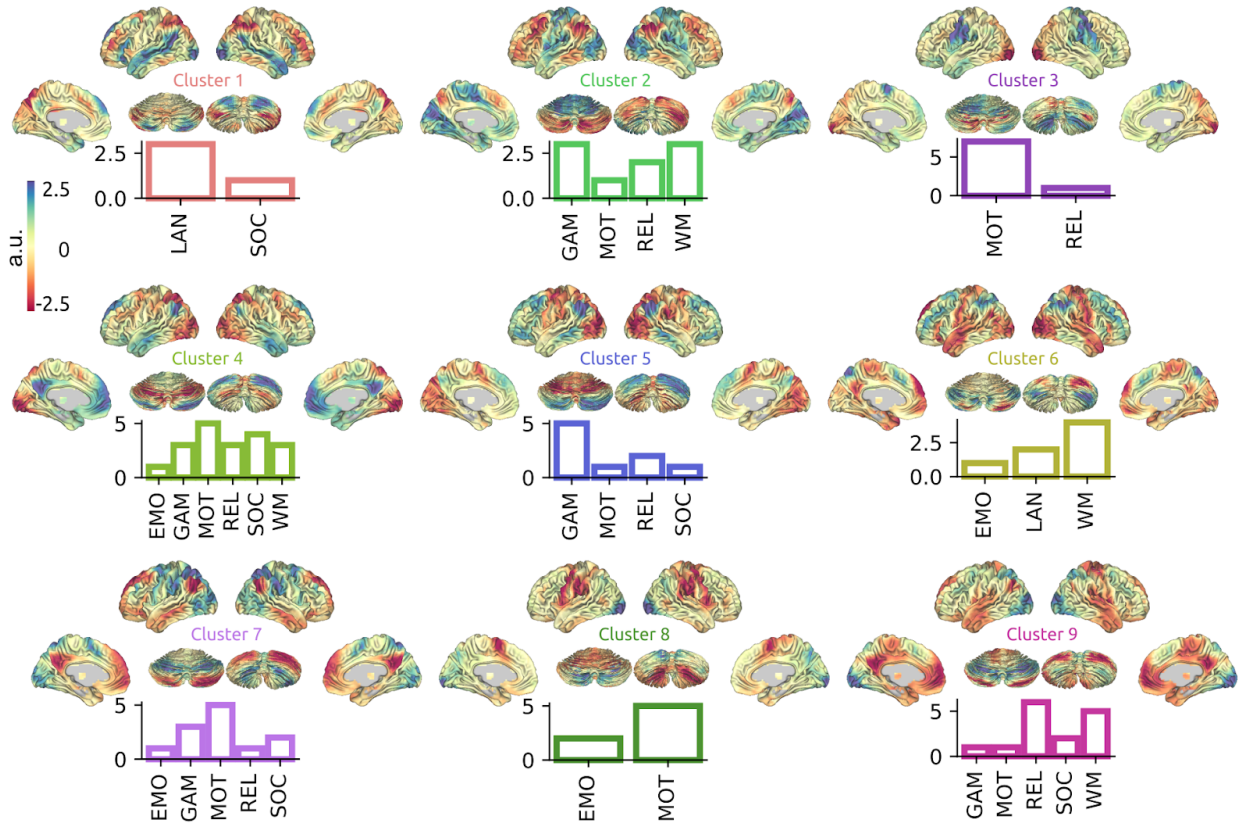


Fig. Sup 27. Spatial meta-subtypes maps

Chapitre 3

Discussion

Le but principal de cette thèse était d'étudier la variabilité et l'universalité du comportement humain à travers les lentilles de l'imagerie cérébrale et de la génétique quantitative. Afin d'examiner l'existence d'un phénotype cérébral qui soit à la fois universel et unique, l'hypothèse voulait que si un tel phénotype existait il devait présenter un ancrage à la fois génétique, environnemental et comportemental.

Dans cette thèse, une étape préalable était nécessaire pour analyser les données en imagerie cérébrale, c'est-à-dire la mise au point d'une méthode pour parvenir à réduire la variabilité dans les données liée à la qualité du prétraitement en IRMf. Dans le chapitre 3, le but de l'étude était de réduire cette variabilité à travers la conception d'un protocole de QC simple et concis. Le protocole développé dans cette étude ne requiert pas de connaissance préalable de l'anatomie du cerveau ni une expertise dans le domaine du QC des images en IRMf. Les résultats montrent qu'indépendamment du niveau d'expertise en QC, les évaluateurs s'accordent sur la qualité de la majorité des images évaluées. Cette étape a permis d'éliminer la variabilité dans les données liée à la méthodologie.

Il s'agissait ensuite d'explorer la variabilité des activations cérébrales parmi un groupe important de sujets. En utilisant 7 tâches cognitives et motrices exécutées dans le scanner d'IRM, j'ai pu estimer la variabilité et l'universalité des activations cérébrales au sein d'un échantillon de 862 jeunes adultes. J'ai regroupé les individus selon la ressemblance de leurs patrons d'activité cérébrale afin de former des sous-groupes homogènes (sous-types), et ce pour chaque tâche. Ainsi, lors d'une même épreuve, les sujets qui forment chaque sous-type neurofonctionnel démontrent un schéma d'activation cérébrale similaire entre eux, mais différent du reste des individus formant les autres sous-types (chapitre 4). Cette technique a permis de rendre compte à la fois de la variabilité et de l'universalité de certains corrélats

neurofonctionnels du comportement humain.

L'étape suivante consistait à étudier l'effet des facteurs génétiques sur les sous-types, ainsi que leurs influences sur le comportement, afin de valider le statut de phénotype intermédiaire de ces sous-types neurofonctionnels entre gènes et comportement. Les résultats montrent que les sous-types sont bel et bien associés à des mesures comportementales telles que la performance à un ensemble d'épreuves socio-cognitives et de personnalité (chapitre 4). Sur le plan génétique, les résultats montrent que les sous-types ainsi que les mesures comportementales affichent une héritabilité variant de modérée à élevée, et que ces derniers présentent aussi une relation pléiotropique (chapitre 4). Ceci indique que certains sous-types d'activation cérébrale et certaines mesures comportementales sont partiellement influencés par les mêmes gènes.

Enfin, la dernière étape consistait à vérifier l'existence d'un lien génétique entre sous-types neurofonctionnels appartenant à différents domaines socio-cognitifs. Ainsi, j'ai examiné la présence de pléiotropie entre les sous-types lors de diverses épreuves psychologiques Les résultats démontrent une pléiotropie élevée intra et inter tâches (chapitre 5) et laisse apparaître deux grands modules d'influence génétique. Plus précisément, deux importants modules pléiotropiques (meta-sous-types) partagent des influences génétiques antagonistes et présentent aussi une association avec les épreuves comportementales. Ce résultat démontre qu'il y aurait deux grands modes de variation génétique à travers les tâches qui sépareraient l'échantillon à l'étude, de sorte qu'un individu qui appartient au premier mode serait associé positivement à des traits comportementaux qualifiés dans un spectre positif. À l'inverse, un individu qui appartient au deuxième mode de variation génétique présenterait une association positive avec des traits comportementaux qualifiés dans un spectre négatif.

La contribution globale de cette thèse est de démontrer qu'il est possible de réunir variabilité et universalité dans un même phénotype possédant à la fois un ancrage génétique, environnemental et comportemental. Les contributions spécifiques sont les suivantes :

Réduction de la variabilité liée au prétraitement des données en IRMf

Dans le chapitre 3, L'apport principal de l'étude est une démonstration de l'efficacité d'un protocole simple permettant de détecter une multitude de problèmes liés à l'acquisition et au prétraitement des images en IRMf. Certaines auteurs ont proposé des protocoles similaires tels que MRIQC (Esteban et al., 2017), fMRIPrep (Esteban et al., 2018) ou

CONN (Whitfield-Gabrieli and Nieto-Castanon, 2012), mais notre étude est la seule qui teste la validité du protocole en estimant l'accord au sein d'un large éventail d'évaluateurs.

Un autre apport de cette thèse réside dans la démonstration de la possibilité de faire du QC en IRMf à travers des plateformes de science participative (Simpson et al., 2014). Les évaluateurs non experts de notre étude sont des volontaires de la plateforme Zooniverse (noa). Une étude de Keshavan et collègues (Keshavan et al., 2018a) en QC d'images en IRM a tenté une pareille initiative de science participative avec succès. L'originalité de notre approche relativement à cette étude consiste en l'utilisation de volontaires sans aucun bagage neuroscientifique, alors que dans l'étude de Keshavan et al. 2018a les évaluateurs font partie de la communauté de neuroimagerie.

Enfin, le protocole développé dans cette thèse est devenu le standard dans le laboratoire de SIMEXP pour chaque donnée prétraitée (Urchs et al., 2018). Il a ainsi permis de détecter des anomalies liées au prétraitement des données en IRMf et de les corriger afin de réduire la variabilité induite dans les images, qui nuirait considérablement aux analyses subséquentes.

Les sous-types d'activation cérébrale ont un ancrage génétique et associé avec le comportement.

Dans le chapitre 3, le but était de valider le statut des sous-types d'activation cérébrale en tant que phénotype intermédiaire entre gènes et comportement. Un phénotype intermédiaire doit avoir une base biologique (être héritable), mais aussi partager des influences génétiques avec le phénotype observable (Crosbie et al., 2008; Vrieze et al., 2012). Les résultats de cette partie de la thèse montrent une héritabilité élevée à la fois de certains sous-types et des mesures comportementales. Ainsi, les sous-types neurofonctionnels peuvent être considérés comme des phénotypes intermédiaires entre les gènes et le comportement observable.

Il est important de noter que chacun des participants à l'étude ont exécuté exactement les mêmes tâches pendant les mêmes durées expérimentales. Dans une étude classique en IRMf de tâche, les analyses de groupe se basent sur la prémisse que les corrélats d'activation cérébrale d'un groupe d'individus sains, et dans la même tranche d'âge, vont activer les mêmes régions cérébrales pour une même tâche dans le scanner (universalité des activations) (Gore, 2003; Mumford and Nichols, 2009). Certaines études ont démontré que cette prémisse n'est pas toujours avérée et que la variabilité cérébrale inter-sujets pour une même tâche est assez considérable (Seghier et al., 2007, 2008; Kherif et al., 2009). La méthode de sous-typage utilisée dans cette thèse concilie à la fois universalité et variabilité des activations inter-sujets. La question qui s'est imposée par la suite était la suivante :

quels sont les facteurs qui lient les sujets en un même sous-type, et quels sont ceux qui les différencient du reste des sujets ? En d'autres termes, quelle est l'influence de la variabilité génétique et de l'environnement dans cette organisation en sous-types neurofonctionnels observée dans notre étude ?

L'héritabilité faible à élevée des sous-types démontre une assise biologique à cette division des groupes en sous-types d'activation cérébrale. Les individus qui constituent les sous-types partageraient, en quelque sorte, une signature biologique commune qui aurait des bases à la fois génétiques et environnementales. Les tâches dont les sous-types présentent une héritabilité élevée nous informent quant à l'importance de l'influence de la variabilité génétique dans le façonnement des activités cérébrales lié à ces sous-types, alors que les tâches présentant une faible héritabilité témoignent d'une influence environnementale comme étant la cause principale de la variabilité des sous-types de cette tâche. Dans notre étude, la tâche motrice est un exemple de sous-types à faible héritabilité, alors que la tâche de langage présente une héritabilité élevée. Ces résultats vont dans le sens de l'hypothèse évolutionnaire de l'expansion corticale du cerveau humain, qui porte elle-même à penser à l'existence de bases évolutionnaires à la variabilité fonctionnelle des différentes régions cérébrales (Mueller et al., 2013). Ainsi, les zones cérébrales qui ont connu une expansion récente dans le cours phylogénétique de notre espèce présentent plus de variabilité inter-sujets. Ces régions sont les aires associatives, qui incluent les zones responsables du langage, alors que les zones motrices et visuelles primaires présentent moins de variabilité interindividuelle. La pensée évolutionnaire en biologie soutient que les gènes qui codent un phénotype crucial pour la survie et la reproduction (fitness) de l'individu sont stabilisés à travers l'évolution de l'espèce (Pérusse and Gendreau, 2005). Ainsi, la variabilité génétique d'un tel trait diminue au fil des générations, alors que les caractères dont l'influence sur la fitness est moins importante peuvent accumuler une plus importante variabilité génétique au cours de l'évolution.

Deux grands modes de variation génétiques influencent les mesures comportementales dans notre échantillon.

Dans le chapitre 5, le but était d'explorer les influences génétiques communes à travers les différents domaines cognitifs des 7 tâches à l'étude. L'influence génétique commune, ou pléiotropie, est significativement élevée entre les sous-types de différents domaines cognitifs, mais aussi, l'organisation de la pléiotropie laisse émerger deux grands modules (meta-sous-types) qui expliquent la majorité de la variabilité génétique dans notre échantillon. Ces deux grands modes de variation génétique sont associés aux mesures comportementales de sorte qu'ils divisent les sujets selon leurs scores comportementaux qui s'inscrivent dans un spectre positif-négatif. Les sujets se trouvant dans la partie « positive » du spectre ont des scores

élevés dans les mesures considérées comme des qualités ou des indicateurs personnels positifs, par exemple, des performances élevées aux épreuves de mémoire, aux tâches cognitives et à la mesure de satisfaction dans la vie. De leur côté, les participants se retrouvant dans la partie « négative » du spectre témoignent de scores élevés pour les traits négatifs, par exemple, affects négatifs et troubles intériorisés. Cette division positive-négative a été aussi révélée par une étude de Smith et collègues (Smith et al., 2015) qui ont associé des patrons de connectivité cérébrale avec des mesures comportementales et démographiques. Ces auteurs ont identifié un mode de variation des mesures selon un spectre positif-négatif semblable au nôtre. Ceci dit, il faut être prudent quant à la généralisation de ces résultats à la population, car nos deux études ont porté sur la même base de données du HCP. Les résultats de ce chapitre portent à croire à l'existence d'un substrat neuro-fonctionnel du comportement humain qui serait à la base des différences inter-sujets dans les activations cérébrales observées dans notre étude.

Interdisciplinarité neuroscience, anthropologie et génétique

Les travaux de cette thèse s'inscrivent dans la mouvance multidisciplinaire entre la l'anthropologie, la neuroscience et la génétique. Le but général était de démêler les influences génétiques et environnementales (culturelles) sur les corrélats neurobiologiques liés à des tâches cognitives, émotionnelles, sociales et motrices. L'anthropologie biologique propose qu'une nature humaine universelle, sous-tendue par un génome largement universel, s'exprimant de façon culturellement variable, engendre cette variabilité observable du comportement. La neuroscience donne accès aux sources organiques du comportement, alors que la génétique renvoie aux origines de la variation comportementale. Les comportements à l'étude dans cette thèse ne représentent qu'une infime partie de l'éventail comportemental humain. Néanmoins, il est apparu possible de mettre à jour une origine à la fois génétique et environnementale qui expliquerait en partie la variabilité comportementale au sein de notre espèce.

Références bibliographiques

Brain match. `https://www.zooniverse.org/projects/simexp/brain-match`. Accessed: 2019-12-19.

Brain match. `https://www.zooniverse.org/projects/simexp/brain-match`, 2015a. Accessed: 2018-4-26.

NIH toolbox. `http://www.healthmeasures.net/explore-measurement-systems/nih-toolbox`, 2015b. Accessed: 2019-10-2.

Nuria Y AbdulSabur, Yisheng Xu, Siyuan Liu, Ho Ming Chow, Miranda Baxter, Jessica Carson, and Allen R Braun. Neural correlates and network connectivity underlying narrative production and comprehension: a combined fMRI and PET study. *Cortex*, 57:107–127, August 2014.

Y Ad-Dab'bagh, O Lyttelton, J S Muehlboeck, C Lepage, D Einarson, K Mok, O Ivanov, R D Vincent, J Lerch, E Fombonne, and Others. The CIVET image-processing environment: a fully automated comprehensive pipeline for anatomical neuroimaging research. In *Proceedings of the 12th annual meeting of the organization for human brain mapping*, page 2266, 2006.

Fidel Alfaro-Almagro, Mark Jenkinson, Neal K Bangerter, Jesper L R Andersson, Ludovica Griffanti, Gwenaëlle Douaud, Stamatios N Sotiropoulos, Saad Jbabdi, Moises Hernandez-Fernandez, Emmanuel Vallee, Diego Vidaurre, Matthew Webster, Paul McCarthy, Christopher Rorden, Alessandro Daducci, Daniel C Alexander, Hui Zhang, Iulius Dragonu, Paul M Matthews, Karla L Miller, and Stephen M Smith. Image processing and quality control for the first 10,000 brain imaging datasets from UK biobank. *Neuroimage*, 166: 400–424, February 2018.

L Almasy and J Blangero. Multipoint quantitative-trait linkage analysis in general pedigrees. *Am. J. Hum. Genet.*, 62(5):1198–1211, May 1998.

Laura Almasy, Thomas D Dyer, and John Blangero. Bivariate quantitative trait linkage analysis: Pleiotropy versus co-incident linkages, 1997.

John Ashburner. A fast diffeomorphic image registration algorithm. *Neuroimage*, 38(1): 95–113, October 2007.

Brian B Avants, Nick Tustison, and Gang Song. Advanced normalization tools (ANTS). *Insight J.*, 2:1–35, 2009.

Abbas Babajani-Feremi. Neural mechanism underling comprehension of narrative speech and its heritability: Study in a large population. *Brain Topogr.*, 30(5):592–609, September 2017.

Lea L Backhausen, Megan M Herting, Judith Buse, Veit Roessner, Michael N Smolka, and Nora C Vetter. Quality control of structural MRI images applied using FreeSurfer-A Hands-On workflow to rate motion artifacts. *Front. Neurosci.*, 10:558, December 2016.

Amanpreet Badhwar, G Peggy McFall, Shraddha Sapkota, Sandra E Black, Howard Chertkow, Simon Duchesne, Mario Masellis, Liang Li, Roger A Dixon, and Pierre Bellec. A multiomics approach to heterogeneity in alzheimer's disease: focused review and roadmap. *Brain*, December 2019a.

Amanpreet Badhwar, G Peggy McFall, Shraddha Sapkota, Sandra E Black, Howard Chertkow, Simon Duchesne, Mario Masellis, Liang Li, Roger Dixon, and Pierre Bellec. A multiomics approach to heterogeneity in alzheimer's disease: Focused review and roadmap. *medRxiv*, page 19008615, October 2019b.

Susan Baldwin. Compute canada: Advancing computational research. *J. Phys. Conf. Ser.*, 341(1):012001, February 2012.

Deanna M Barch, Gregory C Burgess, Michael P Harms, Steven E Petersen, Bradley L Schlaggar, Maurizio Corbetta, Matthew F Glasser, Sandra Curtiss, Sachin Dixit, Cindy Feldt, Dan Nolan, Edward Bryant, Tucker Hartley, Owen Footer, James M Bjork, Russ Poldrack, Steve Smith, Heidi Johansen-Berg, Abraham Z Snyder, David C Van Essen, and WU-Minn HCP Consortium. Function in the human connectome: task-fMRI and individual differences in behavior. *Neuroimage*, 80:169–189, October 2013.

P Bellec, F M Carbonell, V Perlbarg, C Lepage, O Lyttelton, V Fonov, A Janke, J Tohka, and A C Evans. A neuroimaging analysis kit for matlab and octave. In *Proceedings of the 17th International Conference on Functional Mapping of the Human Brain*, pages 2735–2746. Quebec, 2011.

Pierre Bellec, Sébastien Lavoie-Courchesne, Phil Dickinson, Jason P. Lerch, Alex P. Zijdenbos, and Alan C. Evans. The pipeline system for octave and matlab (PSOM): a lightweight scripting framework and execution engine for scientific workflows. *Front. Neuroinform.*, 6, April 2012.

Pierre Bellec, Carlton Chu, François Chouinard-Decorte, Yassine Benhajali, Daniel S Margulies, and R Cameron Craddock. The neuro bureau ADHD-200 preprocessed repository. *Neuroimage*, 144(Pt B):275–286, January 2017.

Yassine Benhajali, Amanpreet Badhwar, Helen Spiers, Sebastian Urchs, Jonathan Armoza, Thomas Ong, Daniel Pérusse, and Pierre Bellec. A standardized protocol for efficient and

reliable quality control of brain registration in functional MRI studies. *Front. Neuroinform.*, 14:7, 2020a.

Yassine Benhajali, Amanpreet Badhwar, Sebastian Urchs, Clara Moreau, Francois Chouinard-Decorte, Uku Vainik, Perrine Ferré, Pierre Orban, Daniel Pérusse, and Pierre Bellec. Subtypes of brain activation are heritable and genetically linked with behavior in the human connectome project sample. June 2020b.

Maxwell Bertolero, Graham Baum, Theodore Satterthwaite, and Danielle Bassett. Brain connectivity is modularly represented in the genome, 2018.

Maxwell A Bertolero, B T Thomas Yeo, and Mark D'Esposito. The modular and integrative functional architecture of the human brain. *Proc. Natl. Acad. Sci. U. S. A.*, 112(49): E6798–807, December 2015.

Maxwell A Bertolero, Ann Sizemore Blevins, Graham L Baum, Ruben C Gur, Raquel E Gur, David R Roalf, Theodore D Satterthwaite, and Danielle S Bassett. The human brain's network architecture is genetically encoded by modular pleiotropy. May 2019.

Richard F Betzel, Maxwell A Bertolero, and Danielle S Bassett. Non-assortative community structure in resting and task-evoked functional brain networks. June 2018.

Bharat B Biswal, Maarten Mennes, Xi-Nian Zuo, Suril Gohel, Clare Kelly, Steve M Smith, Christian F Beckmann, Jonathan S Adelstein, Randy L Buckner, Stan Colcombe, Anne-Marie Dogonowski, Monique Ernst, Damien Fair, Michelle Hampson, Matthew J Hoptman, James S Hyde, Vesa J Kiviniemi, Rolf Kötter, Shi-Jiang Li, Ching-Po Lin, Mark J Lowe, Clare Mackay, David J Madden, Kristoffer H Madsen, Daniel S Margulies, Helen S Mayberg, Katie McMahon, Christopher S Monk, Stewart H Mostofsky, Bonnie J Nagel, James J Pekar, Scott J Peltier, Steven E Petersen, Valentin Riedl, Serge A R B Rombouts, Bart Rypma, Bradley L Schlaggar, Sein Schmidt, Rachael D Seidler, Greg J Siegle, Christian Sorg, Gao-Jun Teng, Juha Veijola, Arno Villringer, Martin Walter, Lihong Wang, Xu-Chu Weng, Susan Whitfield-Gabrieli, Peter Williamson, Christian Windischberger, Yu-Feng Zang, Hong-Ying Zhang, F Xavier Castellanos, and Michael P Milham. Toward discovery science of human brain function. *Proc. Natl. Acad. Sci. U. S. A.*, 107(10): 4734–4739, March 2010.

Henriette Bloch. *Grand dictionnaire de la psychologie*. Larousse, 1999.

Gabriëlla A M Blokland, Katie L McMahon, Jan Hoffman, Gu Zhu, Matthew Meredith, Nicholas G Martin, Paul M Thompson, Greig I de Zubicaray, and Margaret J Wright. Quantifying the heritability of task-related brain activation and performance during the n-back working memory task: a twin fMRI study. *Biol. Psychol.*, 79(1):70–79, September 2008.

Gabriëlla A M Blokland, Katie L McMahon, Paul M Thompson, Nicholas G Martin, Greig I de Zubicaray, and Margaret J Wright. Heritability of working memory brain activation. *J. Neurosci.*, 31(30):10882–10890, July 2011.

Gabriëlla A M Blokland, Greig I de Zubicaray, Katie L McMahon, and Margaret J Wright. Genetic and environmental influences on neuroimaging phenotypes: a meta-analytical perspective on twin imaging studies. *Twin Res. Hum. Genet.*, 15(3):351–371, June 2012.

Gabriëlla A M Blokland, Angus K Wallace, Narelle K Hansell, Paul M Thompson, Ian B Hickie, Grant W Montgomery, Nicholas G Martin, Katie L McMahon, Greig I de Zubicaray, and Margaret J Wright. Genome-wide association study of working memory brain activation. *Int. J. Psychophysiol.*, 115:98–111, May 2017.

Thomas A W Bolton, Valeria Kebets, Enrico Glerean, Daniela Zöller, Jingwei Li, B T Thomas Yeo, César Caballero-Gaudes, and Dimitri Van De Ville. Agito ergo sum: Correlates of spatio-temporal motion characteristics during fMRI. *Neuroimage*, 209:116433, April 2020.

D I Boomsma. Using multivariate genetic modeling to detect pleiotropic quantitative trait loci. *Behav. Genet.*, 26(2):161–166, March 1996.

Randy L Buckner and Fenna M Krienen. The evolution of distributed association networks in the human brain. *Trends Cogn. Sci.*, 17(12):648–665, December 2013.

Nereida Bueno-Guerra. How to apply the concept of umwelt in the evolutionary study of cognition. *Front. Psychol.*, 9:2001, October 2018.

Katherine S Button, John P A Ioannidis, Claire Mokrysz, Brian A Nosek, Jonathan Flint, Emma S J Robinson, and Marcus R Munafò. Power failure: why small sample size undermines the reliability of neuroscience. *Nat. Rev. Neurosci.*, 14(5):365–376, May 2013.

César Caballero-Gaudes and Richard C Reynolds. Methods for cleaning the BOLD fMRI signal. *Neuroimage*, 154:128–149, July 2017.

Vince D Calhoun, Tor D Wager, Anjali Krishnan, Keri S Rosch, Karen E Seymour, Mary Beth Nebel, Stewart H Mostofsky, Prashanth Nyalakanai, and Kent Kiehl. The impact of T1 versus EPI spatial normalization templates for fMRI data analyses. *Hum. Brain Mapp.*, 38(11):5331–5342, November 2017.

Carolin Cardamone, Kevin Schawinski, Marc Sarzi, Steven P Bamford, Nicola Bennert, C M Urry, Chris Lintott, William C Keel, John Parejko, Robert C Nichol, Daniel Thomas, Dan Andreescu, Phil Murray, M Jordan Raddick, Anže Slosar, Alex Szalay, and Jan VandenBerg. Galaxy zoo green peas: discovery of a class of compact extremely star-forming galaxies. *Mon. Not. R. Astron. Soc.*, 399(3):1191–1205, November 2009.

B J Casey, Tariq Cannonier, May I Conley, Alexandra O Cohen, Deanna M Barch, Mary M Heitzeg, Mary E Soules, Theresa Teslovich, Danielle V Dellarco, Hugh Garavan, Catherine A Orr, Tor D Wager, Marie T Banich, Nicole K Speer, Matthew T Sutherland, Michael C Riedel, Anthony S Dick, James M Bjork, Kathleen M Thomas, Bader Chaarani, Margie H Mejia, Donald J Hagler, Jr, M Daniela Cornejo, Chelsea S Sicat, Michael P Harms, Nico U F Dosenbach, Monica Rosenberg, Eric Earl, Hauke Bartsch, Richard Watts, Jonathan R Polimeni, Joshua M Kuperman, Damien A Fair, Anders M Dale, and ABCD

Imaging Acquisition Workgroup. The adolescent brain cognitive development (ABCD) study: Imaging acquisition across 21 sites. *Dev. Cogn. Neurosci.*, 32:43–54, August 2018.

Bernard Chapais and Daniel Pérusse. Présentation : Origine et évolution du comportement humain. *as*, 12(3):1–11, 1988.

Biqing Chen, Zijian Zhu, Yingying Wang, Xiaohu Ding, Xiaobo Guo, Mingguang He, Wan Fang, Qin Zhou, Shanbi Zhou, Han Lei, Ailong Huang, Tingmei Chen, Dongsheng Ni, Yuping Gu, Jianing Liu, and Yi Rao. Nature vs. nurture in human sociality: multi-level genomic analyses of social conformity. *J. Hum. Genet.*, 63(5):605–619, May 2018.

Gang Chen, Ziad S Saad, Audrey R Nath, Michael S Beauchamp, and Robert W Cox. FMRI group analysis combining effect estimates and their variances. *Neuroimage*, 60(1):747–765, March 2012.

James M Cheverud. A COMPARISON OF GENETIC AND PHENOTYPIC CORRELATIONS. *Evolution*, 42(5):958–968, September 1988.

Jacob Cohen. A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.*, 20(1): 37–46, April 1960.

Jacob Cohen. Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychol. Bull.*, 70(4):213–220, 1968.

Michael W Cole, Danielle S Bassett, Jonathan D Power, Todd S Braver, and Steven E Petersen. Intrinsic and task-evoked network architectures of the human brain. *Neuron*, 83 (1):238–251, July 2014.

D L Collins and A C Evans. Animal: Validation and applications of nonlinear Registration-Based segmentation. *Int. J. Pattern Recognit Artif Intell.*, 11(08):1271–1294, December 1997.

D L Collins, P Neelin, T M Peters, and A C Evans. Automatic 3D intersubject registration of MR volumetric data in standardized talairach space. *J. Comput. Assist. Tomogr.*, 18 (2):192–205, March 1994.

Andrew M Colman. *A Dictionary of Psychology*. Oxford University Press, 2009.

Baptiste Couvy-Duchesne, Gabriëlla A M Blokland, Ian B Hickie, Paul M Thompson, Nicholas G Martin, Greig I de Zubicaray, Katie L McMahon, and Margaret J Wright. Heritability of head motion during resting state functional MRI in 462 healthy twins. *Neuroimage*, 102P2:424–434, August 2014.

Baptiste Couvy-Duchesne, Jane L Ebejer, Nathan A Gillespie, David L Duffy, Ian B Hickie, Paul M Thompson, Nicholas G Martin, Greig I de Zubicaray, Katie L McMahon, Sarah E Medland, and Margaret J Wright. Head motion and Inattention/Hyperactivity share common genetic influences: Implications for fMRI studies of ADHD. *PLoS One*, 11(1): e0146271, January 2016.

Jennifer Crosbie, Daniel Pérusse, Cathy L Barr, and Russell J Schachar. Validating psychiatric endophenotypes: inhibitory control and attention deficit hyperactivity disorder.

Neurosci. Biobehav. Rev., 32(1):40–55, 2008.

Mahsa Dadar, Vladimir S Fonov, D Louis Collins, and Alzheimer's Disease Neuroimaging Initiative. A comparison of publicly available linear MRI stereotaxic registration techniques. *Neuroimage*, 174:191–200, July 2018.

Sean P David, Jennifer J Ware, Isabella M Chu, Pooja D Loftus, Paolo Fusar-Poli, Joaquim Radua, Marcus R Munafò, and John P A Ioannidis. Potential reporting bias in fMRI studies of the brain. *PLoS One*, 8(7):e70104, July 2013.

A E Desjardins, K A Kiehl, and P F Liddle. Removal of confounding effects of global signal in functional MRI analyses. *Neuroimage*, 13(4):751–758, April 2001.

Andrew T Drysdale, Logan Grosenick, Jonathan Downar, Katharine Dunlop, Farrokh Mansouri, Yue Meng, Robert N Fetcho, Benjamin Zebley, Desmond J Oathes, Amit Etkin, Alan F Schatzberg, Keith Sudheimer, Jennifer Keller, Helen S Mayberg, Faith M Gunning, George S Alexopoulos, Michael D Fox, Alvaro Pascual-Leone, Henning U Voss, B J Casey, Marc J Dubin, and Conor Liston. Resting-state connectivity biomarkers define neurophysiological subtypes of depression. *Nat. Med.*, December 2016.

Simon Ducharme, Matthew D Albaugh, Tuong-Vi Nguyen, James J Hudziak, J M Mateos-Pérez, Aurelie Labbe, Alan C Evans, Sherif Karama, and Brain Development Cooperative Group. Trajectories of cortical thickness maturation in normal brain development–the importance of quality control procedures. *Neuroimage*, 125:267–279, January 2016.

Amanda K Easson, Zainab Fatima, and Anthony R McIntosh. Functional connectivity-based subtypes of individuals with and without autism spectrum disorder. *Netw Neurosci*, 3(2): 344–362, February 2019.

Lloyd T Elliott, Kevin Sharp, Fidel Alfaro-Almagro, Sinan Shi, Karla Miller, Gwenaëlle Douaud, Jonathan Marchini, and Stephen Smith. Genome-wide association studies of brain structure and function in the UK biobank. June 2018.

Laura E Engelhardt, Mary Abbe Roe, Jenifer Juranek, Dana DeMaster, K Paige Harden, Elliot M Tucker-Drob, and Jessica A Church. Children's head motion during fMRI tasks is heritable and stable over time. *Dev. Cogn. Neurosci.*, February 2017.

Oscar Esteban, Daniel Birman, Marie Schaer, Oluwasanmi O Koyejo, Russell A Poldrack, and Krzysztof J Gorgolewski. MRIQC: Advancing the automatic prediction of image quality in MRI from unseen sites. *PLoS One*, 12(9):e0184661, September 2017.

Oscar Esteban, Ross Blair, Christopher J Markiewicz, Shoshana L Berleant, Craig Moodie, Feilong Ma, Ayse Ilkay Isik, Asier Erramuzpe, James D Kent, Mathias Goncalves, Elizabeth DuPre, Kevin R Sitek, Daniel E P Gomez, Daniel J Lurie, Zhifang Ye, Taylor Salo, Romain Valabregue, Inge K Amlien, Franz Liem, Nir Jacoby, Russell A Poldrack, and Krzysztof J Gorgolewski. fmriprep, December 2018.

Oscar Esteban, Christopher J Markiewicz, Ross W Blair, Craig A Moodie, A Ilkay Isik, Asier Erramuzpe, James D Kent, Mathias Goncalves, Elizabeth DuPre, Madeleine Snyder,

Hiroyuki Oya, Satrajit S Ghosh, Jessey Wright, Joke Durnez, Russell A Poldrack, and Krzysztof J Gorgolewski. fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nat. Methods*, 16(1):111–116, January 2019.

DS Falconer, Tfc Mackay, and others. Introduction to quantitative genetics (4th edn). *Investig. Genet.*, 1996.

Eva Feredoes and Bradley R Postle. Localization of load sensitivity of working memory storage: Quantitatively and qualitatively discrepant results yielded by single-subject and group-averaged approaches to fMRI group analysis, 2007.

Anders M Fjell, Lars T Westlye, Inge Amlien, Christian K Tamnes, Håkon Grydeland, Andreas Engvig, Thomas Espeseth, Ivar Reinvang, Astri J Lundervold, Arvid Lundervold, and Kristine B Walhovd. High-Expanding cortical regions in human development and evolution are related to higher intellectual abilities. *Cereb. Cortex*, August 2013.

V F Flack, A A Afifi, P A Lachenbruch, and H J A Schouten. Sample size determinations for the two rater kappa statistic. *Psychometrika*, 53(3):321–325, September 1988.

V S Fonov, A C Evans, R C McKinstry, C R Almli, and D L Collins. Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *Neuroimage*, 47:S102, July 2009.

Vladimir Fonov, Pierrick Coupe, Simon Eskildsen, and D Collins. Atrophy-specific MRI brain template for alzheimer's disease and mild cognitive impairment. *Alzheimers. Dement.*, 7 (4, Supplement):S58, July 2011a.

Vladimir Fonov, Alan C Evans, Kelly Botteron, C Robert Almli, Robert C McKinstry, D Louis Collins, and Brain Development Cooperative Group. Unbiased average age-appropriate atlases for pediatric studies. *Neuroimage*, 54(1):313–327, January 2011b.

Vladimir Fonov, Mahsa Dadar, The PREVENT-AD Research Group, and D Louis Collins. Deep learning of quality control for stereotaxic registration of human brain MRI. April 2018.

Chiara Franzoni and Henry Sauermann. Crowd science: The organization of scientific research in open collaborative projects. *Res. Policy*, 43(1):1–20, February 2014.

K J Friston, A P Holmes, K J Worsley, J-P Poline, C D Frith, and R S J Frackowiak. Statistical parametric maps in functional imaging: A general linear approach. *Hum. Brain Mapp.*, 2(4):189–210, October 1994.

Matthias Gamer. irr, January 2012.

Habib Ganjgahi, Anderson M Winkler, David C Glahn, John Blangero, Peter Kochunov, and Thomas E Nichols. Fast and powerful heritability inference for family-based neuroimaging studies. *Neuroimage*, 115:256–268, July 2015.

Clifford Geertz and Others. Thick description: Toward an interpretive theory of culture. *Turning points in qualitative research: Tying knots in a handkerchief*, 3:143–168, 1973.

Alysha Gilmore, Nicholas Buser, and Jamie L Hanson. Variations in structural MRI quality impact measures of brain anatomy: Relations with age and other sociodemographic variables. March 2019.

Federico Giove, Tommaso Gili, Vittorio Iacovella, Emiliano Macaluso, and Bruno Maraviglia. Images-based suppression of unwanted global signals in resting-state functional connectivity studies. *Magn. Reson. Imaging*, 27(8):1058–1064, October 2009.

Matthew F Glasser, Stamatios N Sotiropoulos, J Anthony Wilson, Timothy S Coalson, Bruce Fischl, Jesper L Andersson, Junqian Xu, Saad Jbabdi, Matthew Webster, Jonathan R Polimeni, David C Van Essen, Mark Jenkinson, and WU-Minn HCP Consortium. The minimal preprocessing pipelines for the human connectome project. *Neuroimage*, 80:105–124, October 2013.

Matthew F Glasser, Stephen M Smith, Daniel S Marcus, Jesper L R Andersson, Edward J Auerbach, Timothy E J Behrens, Timothy S Coalson, Michael P Harms, Mark Jenkinson, Steen Moeller, Emma C Robinson, Stamatios N Sotiropoulos, Junqian Xu, Essa Yacoub, Kamil Ugurbil, and David C Van Essen. The human connectome project's neuroimaging approach. *Nat. Neurosci.*, 19(9):1175–1187, August 2016.

John C Gore. Principles and practice of functional MRI of the human brain. *J. Clin. Invest.*, 112(1):4–9, July 2003.

Krzysztof Gorgolewski, Oscar Esteban, Gunnar Schaefer, Brian Wandell, and Russell Poldrack. Openneuro-a free online platform for sharing and analysis of neuroimaging data, 2017. *URL https://ww5. aievolution. com/hbm1701/index. cfm.*

Katrina L Grasby, Neda Jahanshad, Jodie N Painter, Lucía Colodro-Conde, Janita Bralten, Derrek P Hibar, Penelope A Lind, Fabrizio Pizzagalli, Christopher R K Ching, Mary Agnes B McMahon, Natalia Shatokhina, Leo C P Zsembik, Sophia I Thomopoulos, Alyssa H Zhu, Lachlan T Strike, Ingrid Agartz, Saud Alhusaini, Marcio A A Almeida, Dag Alnæs, Inge K Amlien, Micael Andersson, Tyler Ard, Nicola J Armstrong, Allison Ashley-Koch, Joshua R Atkins, Manon Bernard, Rachel M Brouwer, Elizabeth E L Buimer, Robin Bülow, Christian Bürger, Dara M Cannon, Mallar Chakravarty, Qiang Chen, Joshua W Cheung, Baptiste Couvy-Duchesne, Anders M Dale, Shareefa Dalvie, Tânia K de Araujo, Greig I de Zubicaray, Sonja M C de Zwarte, Anouk den Braber, Nhat Trung Doan, Katharina Dohm, Stefan Ehrlich, Hannah-Ruth Engelbrecht, Susanne Erk, Chun Chieh Fan, Iryna O Fedko, Sonya F Foley, Judith M Ford, Masaki Fukunaga, Melanie E Garrett, Tian Ge, Sudheer Giddaluru, Aaron L Goldman, Melissa J Green, Nynke A Groenewold, Dominik Grotegerd, Tiril P Gurholt, Boris A Gutman, Narelle K Hansell, Mathew A Harris, Marc B Harrison, Courtney C Haswell, Michael Hauser, Stefan Herms, Dirk J Heslenfeld, New Fei Ho, David Hoehn, Per Hoffmann, Laurena Holleran, Martine Hoogman, Jouke-Jan Hottenga, Masashi Ikeda, Deborah Janowitz, Iris E Jansen, Tianye Jia, Christiane Jockwitz, Ryota Kanai, Sherif Karama, Dalia Kasperaviciute,

Tobias Kaufmann, Sinead Kelly, Masataka Kikuchi, Marieke Klein, Michael Knapp, Annchen R Knodt, Bernd Krämer, Max Lam, Thomas M Lancaster, Phil H Lee, Tristram A Lett, Lindsay B Lewis, Iscia Lopes-Cendes, Michelle Luciano, Fabio Macciardi, Andre F Marquand, Samuel R Mathias, Tracy R Melzer, Yuri Milaneschi, Nazanin Mirza-Schreiber, Jose C V Moreira, Thomas W Mühleisen, Bertram Müller-Myhsok, Pablo Najt, Soichiro Nakahara, Kwangsik Nho, Loes M Olde Loohuis, Dimitri Papadopoulos Orfanos, John F Pearson, Toni L Pitcher, Benno Pütz, Yann Quidé, Anjanibhargavi Ragothaman, Faisal M Rashid, William R Reay, Ronny Redlich, Céline S Reinbold, Jonathan Repple, Geneviève Richard, Brandalyn C Riedel, Shannon L Risacher, Cristiane S Rocha, Nina Roth Mota, Lauren Salminen, Arvin Saremi, Andrew J Saykin, Fenja Schlag, Lianne Schmaal, Peter R Schofield, Rodrigo Secolin, Chin Yang Shapland, Li Shen, Jean Shin, Elena Shumskaya, Ida E Sønderby, Emma Sprooten, Katherine E Tansey, Alexander Teumer, Anbupalam Thalamuthu, Diana Tordesillas-Gutiérrez, Jessica A Turner, Anne Uhlmann, Costanza Ludovica Vallerga, Dennis van der Meer, Marjolein M J van Donkelaar, Liza van Eijk, Theo G M van Erp, Neeltje E M van Haren, Daan van Rooij, Marie-José van Tol, Jan H Veldink, Ellen Verhoef, Esther Walton, Mingyuan Wang, Yunpeng Wang, Joanna M Wardlaw, Wei Wen, Lars T Westlye, Christopher D Whelan, Stephanie H Witt, Katharina Wittfeld, Christiane Wolf, Thomas Wolfers, Jing Qin Wu, Clarissa L Yasuda, Dario Zaremba, Zuo Zhang, Marcel P Zwiers, Eric Artiges, Amelia A Assareh, Rosa Ayesa-Arriola, Aysenil Belger, Christine L Brandt, Gregory G Brown, Sven Cichon, Joanne E Curran, Gareth E Davies, Franziska Degenhardt, Michelle F Dennis, Bruno Dietsche, Srdjan Djurovic, Colin P Doherty, Ryan Espiritu, Daniel Garijo, Yolanda Gil, Penny A Gowland, Robert C Green, Alexander N Häusler, Walter Heindel, Beng-Choon Ho, Wolfgang U Hoffmann, Florian Holsboer, Georg Homuth, Norbert Hosten, Clifford R Jack, Mihyun Jang, Andreas Jansen, Nathan A Kimbrel, Knut Kolskår, Sanne Koops, Axel Krug, Kelvin O Lim, Jurjen J Luykx, Daniel H Mathalon, Karen A Mather, Venkata S Mattay, Sarah Matthews, Jaqueline Mayoral Van Son, Sarah C McEwen, Ingrid Melle, Derek W Morris, Bryon A Mueller, Matthias Nauck, Jan E Nordvik, Markus M Nöthen, Daniel S O'Leary, Nils Opel, Marie-Laure Paillère Martinot, G Bruce Pike, Adrian Preda, Erin B Quinlan, Paul E Rasser, Varun Ratnakar, Simone Reppermund, Vidar M Steen, Paul A Tooney, Fábio R Torres, Dick J Veltman, James T Voyvodic, Robert Whelan, Tonya White, Hidenaga Yamamori, Hieab H H Adams, Joshua C Bis, Stephanie Debette, Charles Decarli, Myriam Fornage, Vilmundur Gudnason, Edith Hofer, M Arfan Ikram, Lenore Launer, W T Longstreth, Oscar L Lopez, Bernard Mazoyer, Thomas H Mosley, Gennady V Roshchupkin, Claudia L Satizabal, Reinhold Schmidt, Sudha Seshadri, Qiong Yang, Alzheimer's Disease Neuroimaging Initiative¶, CHARGE Consortium¶, EPIGEN Consortium¶, IMAGEN Consortium¶, SYS Consortium¶, Parkinson's Progression Markers Initiative¶,

Marina K M Alvim, David Ames, Tim J Anderson, Ole A Andreassen, Alejandro Arias-Vasquez, Mark E Bastin, Bernhard T Baune, Jean C Beckham, John Blangero, Dorret I Boomsma, Henry Brodaty, Han G Brunner, Randy L Buckner, Jan K Buitelaar, Juan R Bustillo, Wiepke Cahn, Murray J Cairns, Vince Calhoun, Vaughan J Carr, Xavier Caseras, Svenja Caspers, Gianpiero L Cavalleri, Fernando Cendes, Aiden Corvin, Benedicto Crespo-Facorro, John C Dalrymple-Alford, Udo Dannlowski, Eco J C de Geus, Ian J Deary, Norman Delanty, Chantal Depondt, Sylvane Desrivières, Gary Donohoe, Thomas Espeseth, Guillén Fernández, Simon E Fisher, Herta Flor, Andreas J Forstner, Clyde Francks, Barbara Franke, David C Glahn, Randy L Gollub, Hans J Grabe, Oliver Gruber, Asta K Håberg, Ahmad R Hariri, Catharina A Hartman, Ryota Hashimoto, Andreas Heinz, Frans A Henskens, Manon H J Hillegers, Pieter J Hoekstra, Avram J Holmes, L Elliot Hong, William D Hopkins, Hilleke E Hulshoff Pol, Terry L Jernigan, Erik G Jönsson, René S Kahn, Martin A Kennedy, Tilo T J Kircher, Peter Kochunov, John B J Kwok, Stephanie Le Hellard, Carmel M Loughland, Nicholas G Martin, Jean-Luc Martinot, Colm McDonald, Katie L McMahon, Andreas Meyer-Lindenberg, Patricia T Michie, Rajendra A Morey, Bryan Mowry, Lars Nyberg, Jaap Oosterlaan, Roel A Ophoff, Christos Pantelis, Tomas Paus, Zdenka Pausova, Brenda W J, Tinca J C Polderman, Danielle Posthuma, Marcella Rietschel, Joshua L Roffman, Laura M Rowland, Perminder S Sachdev, Philipp G Sämann, Ulrich Schall, Gunter Schumann, Rodney J Scott, Kang Sim, Sanjay M Sisodiya, Jordan W Smoller, Iris E Sommer, Beate St Pourcain, Dan J Stein, Arthur W Toga, Julian N Trollor, Nic J A Van der Wee, Dennis van 't Ent, Henry Völzke, Henrik Walter, Bernd Weber, Daniel R Weinberger, Margaret J Wright, Juan Zhou, Jason L Stein, Paul M Thompson, Sarah E Medland, and Enhancing NeuroImaging Genetics through Meta-Analysis Consortium (ENIGMA)—Genetics working group. The genetic architecture of the human cerebral cortex. *Science*, 367(6484), March 2020.

W Grodd, E Hülsmann, M Lotze, D Wildgruber, and M Erb. Sensorimotor mapping of the human cerebellum: fMRI evidence of somatotopic organization. *Hum. Brain Mapp.*, 13 (2):55–73, June 2001.

C Habas. Physiological basis of functional MRI. *J. Radiol.*, 83(11):1737–1741, 2002.

Yanting Han and Ralph Adolphs. Estimating the heritability of psychological measures in the human connectome project dataset. *bioRxiv*, 2019.

Uri Hasson, Ohad Landesman, Barbara Knappmeyer, Ignacio Vallines, Nava Rubin, and David J Heeger. Neurocinematics: The neuroscience of film. *Curr. Popul. Rep. Popul. Estim. Proj.*, 2(1):1–26, August 2008.

Colin S Hawco, Erin W Dickie, Grace Jacobs, Zafiris J Daskalakis, and Aristotle N Voineskos. Moving beyond the mean: Subgroups and dimensions of brain activity and cognitive performance across domains. March 2020.

R Heun, F Jessen, U Klose, M Erb, D Granath, N Freymann, and W Grodd. Interindividual variation of cerebral activation during encoding and retrieval of words. *Eur. Psychiatry*, 15(8):470–479, December 2000.

Gregory Hickok and David Poeppel. The cortical organization of speech processing. *Nat. Rev. Neurosci.*, 8(5):393–402, May 2007.

Jason Hill, Terrie Inder, Jeffrey Neil, Donna Dierker, John Harwell, and David Van Essen. Similar patterns of cortical expansion during human development and evolution. *Proc. Natl. Acad. Sci. U. S. A.*, 107(29):13135–13140, July 2010.

Michael R Hodge, William Horton, Timothy Brown, Rick Herrick, Timothy Olsen, Michael E Hileman, Michael McKay, Kevin A Archie, Eileen Cler, Michael P Harms, Gregory C Burgess, Matthew F Glasser, Jennifer S Elam, Sandra W Curtiss, Deanna M Barch, Robert Oostenveld, Linda J Larson-Prior, Kamil Ugurbil, David C Van Essen, and Daniel S Marcus. ConnectomeDB-Sharing human brain connectivity data. *Neuroimage*, 124(Pt B): 1102–1107, January 2016.

J Howe. The rise of crowdsourcing. `https://sistemas-humano-computacionais.wdfiles.com/local--files/capitulo%3Aredes-sociais/Howe_The_Rise_of_Crowdsourcing.pdf`, 2006. Accessed: 2019-5-1.

Andrew Janke and Vladimir S Fonov. Register, 2012.

Mark Jenkinson, Christian F Beckmann, Timothy E J Behrens, Mark W Woolrich, and Stephen M Smith. FSL. *Neuroimage*, 62(2):782–790, August 2012.

Ryota Kanai and Geraint Rees. The structural basis of inter-individual differences in human behaviour and cognition. *Nat. Rev. Neurosci.*, 12(4):231–242, April 2011.

Rajan Kashyap, Sagarika Bhattacharjee, B T Thomas Yeo, and S H Annabel Chen. Maximizing dissimilarity in resting state detects heterogeneous subtypes in healthy population associated with high substance use and problems in antisocial personality. *Hum. Brain Mapp.*, November 2019.

Anisha Keshavan, Esha Datta, Ian M McDonough, Christopher R Madan, Kesshi Jordan, and Roland G Henry. Mindcontrol: A web application for brain segmentation quality control. *Neuroimage*, 170:365–372, April 2018a.

Anisha Keshavan, Jason D Yeatman, and Ariel Rokem. Combining citizen science and deep learning to amplify expertise in neuroimaging. July 2018b.

Firas Khatib, Frank DiMaio, Foldit Contenders Group, Foldit Void Crushers Group, Seth Cooper, Maciej Kazmierczyk, Miroslaw Gilski, Szymon Krzywda, Helena Zabranska, Iva Pichova, James Thompson, Zoran Popović, Mariusz Jaskolski, and David Baker. Crystal structure of a monomeric retroviral protease solved by protein folding game players. *Nat. Struct. Mol. Biol.*, 18(10):1175–1177, September 2011.

Ferath Kherif, Jean-Baptiste Poline, Sébastien Mériaux, Habib Benali, Guillaume Flandin, and Matthew Brett. Group analysis in functional neuroimaging: selecting subjects using

similarity measures. *Neuroimage*, 20(4):2197–2208, December 2003.

Ferath Kherif, Goulven Josse, Mohamed L Seghier, and Cathy J Price. The main sources of intersubject variability in neuronal activation for reading aloud. *J. Cogn. Neurosci.*, 21 (4):654–668, April 2009.

Jinseop S Kim, Matthew J Greene, Aleksandar Zlateski, Kisuk Lee, Mark Richardson, Srinivas C Turaga, Michael Purcaro, Matthew Balkam, Amy Robinson, Bardia F Behabadi, Michael Campos, Winfried Denk, H Sebastian Seung, and EyeWirers. Space-time wiring specificity supports direction selectivity in the retina. *Nature*, 509(7500):331–336, May 2014.

Brenda A Kirchhoff and Randy L Buckner. Functional-anatomic correlates of individual differences in memory. *Neuron*, 51(2):263–274, July 2006.

Eduard T Klapwijk, Ferdi van de Kamp, Mara van der Meulen, Sabine Peters, and Lara M Wierenga. Qoala-T: A supervised-learning tool for quality control of FreeSurfer segmented MRI data. *Neuroimage*, January 2019.

Kalevi Kull. Umwelt and modelling. *The Routledge companion to semiotics*, 43:56, 2010.

J R Landis and G G Koch. The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159–174, March 1977.

Yann Le Guen, Marie Amalric, Philippe Pinel, Christophe Pallier, and Vincent Frouin. Shared genetic aetiology between cognitive performance and brain activations in language and math tasks. *Sci. Rep.*, 8(1):17624, December 2018.

M Lebreton and S Palminteri. Revisiting the assessment of inter-individual differences in fMRI activations-behavior relationships. *bioRxiv*, 2016.

C J Lintott, K Schawinski, W Keel, and others. Galaxy Zoo:'Hanny's voorwerp', a quasar light echo? *Mon. Not. R. Astron. Soc.*, 2009.

Alberto Llera, Thomas Wolfers, Peter Mulders, and Christian F Beckmann. Inter-individual differences in human brain structure and morphology link to variation in demographics and behavior. *Elife*, 8, July 2019.

Nikos K Logothetis and Brian A Wandell. Interpreting the BOLD signal. *Annu. Rev. Physiol.*, 66:735–769, 2004.

Weizhao Lu, Kejiang Dong, Dong Cui, Qing Jiao, and Jianfeng Qiu. Quality assurance of human functional magnetic resonance imaging: a literature review. *Quant. Imaging Med. Surg.*, 9(6):1147–1162, June 2019.

Sonia J Lupien, Bruce S McEwen, Megan R Gunnar, and Christine Heim. Effects of stress throughout the lifespan on the brain, behaviour and cognition. *Nat. Rev. Neurosci.*, 10 (6):434–445, June 2009.

W C Machielsen, S A Rombouts, F Barkhof, P Scheltens, and M P Witter. FMRI of visual encoding: reproducibility of activation. *Hum. Brain Mapp.*, 9(3):156–164, March 2000.

A E Maxwell, P B C Fenwick, G W Fenton, and J Dollimore. Reading ability and brain function: a simple statistical model, 1974.

Andrew R Mayer, David Ruhl, Flannery Merideth, Josef Ling, Faith M Hanlon, Juan Bustillo, and Jose Cañive. Functional imaging of the hemodynamic sensory gating response in schizophrenia. *Hum. Brain Mapp.*, 34(9):2302–2312, September 2013.

D J McGonigle, A M Howseman, B S Athwal, K J Friston, R S Frackowiak, and A P Holmes. Variability in fMRI: an examination of intersession differences. *Neuroimage*, 11(6 Pt 1): 708–734, June 2000.

Maarten Mennes, Bharat B Biswal, F Xavier Castellanos, and Michael P Milham. Making data sharing work: the FCP/INDI experience. *Neuroimage*, 82:683–691, November 2013.

Michael P Milham, R Cameron Craddock, Jake J Son, Michael Fleischmann, Jon Clucas, Helen Xu, Bonhwang Koo, Anirudh Krishnakumar, Bharat B Biswal, F Xavier Castellanos, Stan Colcombe, Adriana Di Martino, Xi-Nian Zuo, and Arno Klein. Assessment of the impact of shared brain imaging data on the scientific literature. *Nat. Commun.*, 9(1):2818, July 2018.

Michael B Miller, John Darrell Van Horn, George L Wolford, Todd C Handy, Monica Valsangkar-Smyth, Souheil Inati, Scott Grafton, and Michael S Gazzaniga. Extensive individual differences in brain activations associated with episodic retrieval are reliable over time. *J. Cogn. Neurosci.*, 14(8):1200–1214, November 2002.

Michael B Miller, Christa-Lynn Donovan, John D Van Horn, Elaine German, Peter Sokol-Hessner, and George L Wolford. Unique and persistent individual patterns of brain activity across different memory retrieval tasks. *Neuroimage*, 48(3):625–635, November 2009.

Michael B Miller, Christa-Lynn Donovan, Craig M Bennett, Elissa M Aminoff, and Richard E Mayer. Individual differences in cognitive style and strategy predict similarities in the patterns of brain activity between individuals. *Neuroimage*, 59(1):83–93, January 2012.

Sophia Mueller, Danhong Wang, Michael D Fox, B T Thomas Yeo, Jorge Sepulcre, Mert R Sabuncu, Rebecca Shafee, Jie Lu, and Hesheng Liu. Individual variability in functional connectivity architecture of the human brain. *Neuron*, 77(3):586–595, February 2013.

J P Mugler and J R Brookeman. Three-dimensional magnetization-prepared rapid gradient-echo imaging (3d MP RAGE). *Magn. Reson. Med.*, 1990.

Jeanette A Mumford and Thomas Nichols. Simple group fMRI modeling and inference. *Neuroimage*, 47(4):1469–1475, October 2009.

M Neale, L Cardon, and North Atlantic Treaty Organization Scientific Affairs Division. *Methodology for Genetic Studies of Twins and Families*. D]: [Nato ASI series. Springer, 1992.

T E Nichols, S Das, S B Eickhoff, A C Evans, T Glatard, M Hanke, N Kriegeskorte, M P Milham, R A Poldrack, J B Poline, E Proal, B Thirion, D C Essen Van, T White, and B T Yeo. Best practices in data analysis and sharing in neuroimaging using MRI. *Nat.*

Neurosci., 20(3):299–303, February 2017.

Kate Brody Nooner, Stanley J Colcombe, Russell H Tobe, Maarten Mennes, Melissa M Benedict, Alexis L Moreno, Laura J Panek, Shaquanna Brown, Stephen T Zavitz, Qingyang Li, Sharad Sikka, David Gutman, Saroja Bangaru, Rochelle Tziona Schlachter, Stephanie M Kamiel, Ayesha R Anwar, Caitlin M Hinz, Michelle S Kaplan, Anna B Rachlin, Samantha Adelsberg, Brian Cheung, Ranjit Khanuja, Chaogan Yan, Cameron C Craddock, Vincent Calhoun, William Courtney, Margaret King, Dylan Wood, Christine L Cox, A M Clare Kelly, Adriana Di Martino, Eva Petkova, Philip T Reiss, Nancy Duan, Dawn Thomsen, Bharat Biswal, Barbara Coffey, Matthew J Hoptman, Daniel C Javitt, Nunzio Pomara, John J Sidtis, Harold S Koplewicz, Francisco Xavier Castellanos, Bennett L Leventhal, and Michael P Milham. The NKI-Rockland sample: A model for accelerating the pace of discovery science in psychiatry. *Front. Neurosci.*, 6:152, October 2012.

Annalise B Paaby and Matthew V Rockman. The many faces of pleiotropy. *Trends Genet.*, 29(2):66–73, February 2013.

Raja Parasuraman and Yang Jiang. Individual differences in cognition, affect, and performance: behavioral, neuroimaging, and molecular genetic approaches. *Neuroimage*, 59(1): 70–82, January 2012.

Jong Geun Park and Chulhee Lee. Skull stripping based on region growing for magnetic resonance brain images. *Neuroimage*, 47(4):1394–1407, October 2009.

Alice Patania, Pierluigi Selvaggi, Mattia Veronese, Ottavia Dipasquale, Paul Expert, and Giovanni Petri. Topological gene expression networks recapitulate brain anatomy and function. *Netw Neurosci*, 3(3):744–762, July 2019.

D Paul. A double-edged sword. *Nature*, 405(6786):515, June 2000.

W Penny and A Holmes. Random effects analysis. *Statistical parametric mapping: The analysis of functional brain images*, pages 156–165, 2007.

D Pérusse and P L Gendreau. Genetics and the development of aggression. *Developmental origins of aggression*, 2005.

Philippe Pinel and Stanislas Dehaene. Genetic and environmental contributions to brain activation during calculation. *Neuroimage*, 81:306–316, November 2013.

Ricardo A Pizarro, Xi Cheng, Alan Barnett, Herve Lemaitre, Beth A Verchinski, Aaron L Goldman, Ena Xiao, Qian Luo, Karen F Berman, Joseph H Callicott, Daniel R Weinberger, and Venkata S Mattay. Automated quality assessment of structural magnetic resonance brain images based on a supervised machine learning algorithm. *Front. Neuroinform.*, 10: 52, December 2016.

Tinca J C Polderman, Beben Benyamin, Christiaan A de Leeuw, Patrick F Sullivan, Arjen van Bochoven, Peter M Visscher, and Danielle Posthuma. Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat. Genet.*, 47(7):702–709, July 2015.

Russell A Poldrack and Krzysztof J Gorgolewski. Making big data open: data sharing in neuroimaging. *Nat. Neurosci.*, 17(11):1510–1517, November 2014.

Jonathan D Power, Kelly A Barnes, Abraham Z Snyder, Bradley L Schlaggar, and Steven E Petersen. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *Neuroimage*, 59(3):2142–2154, February 2012.

Cathy J Price. A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *Neuroimage*, 62(2):816–847, August 2012.

Li-Lin Rao, Yuan Zhou, Dang Zheng, Liu-Qing Yang, and Shu Li. Genetic contribution to variation in risk taking: A functional MRI twin study of the balloon analogue risk task. *Psychol. Sci.*, 29(10):1679–1691, October 2018.

Jonas Richiardi, Andre Altmann, Anna-Clare Milazzo, Catie Chang, M Mallar Chakravarty, Tobias Banaschewski, Gareth J Barker, Arun L W Bokde, Uli Bromberg, Christian Büchel, Patricia Conrod, Mira Fauth-Bühler, Herta Flor, Vincent Frouin, Jürgen Gallinat, Hugh Garavan, Penny Gowland, Andreas Heinz, Hervé Lemaître, Karl F Mann, Jean-Luc Martinot, Frauke Nees, Tomáš Paus, Zdenka Pausova, Marcella Rietschel, Trevor W Robbins, Michael N Smolka, Rainer Spanagel, Andreas Ströhle, Gunter Schumann, Mike Hawrylycz, Jean-Baptiste Poline, Michael D Greicius, and IMAGEN consortium. BRAIN NETWORKS. correlated gene expression supports synchronous activity in brain networks. *Science*, 348(6240):1241–1244, June 2015.

Chris Rorden. MRIcron, 2014.

Adon F G Rosen, David R Roalf, Kosha Ruparel, Jason Blake, Kevin Seelaus, Lakshmi P Villa, Rastko Ciric, Philip A Cook, Christos Davatzikos, Mark A Elliott, Angel Garcia de La Garza, Efstathios D Gennatas, Megan Quarmley, J Eric Schmitt, Russell T Shinohara, M Dylan Tisdall, R Cameron Craddock, Raquel E Gur, Ruben C Gur, and Theodore D Satterthwaite. Quantitative assessment of structural image quality. *Neuroimage*, 169: 407–418, April 2018.

Emi Saliasi, Linda Geerligs, Monicque M Lorist, and Natasha M Maurits. Neural correlates associated with successful working memory performance in older adults as revealed by spatial ICA. *PLoS One*, 9(6):e99250, June 2014.

Mohamed L Seghier and Cathy J Price. Dissociating functional brain networks by decoding the between-subject variability. *Neuroimage*, 45(2):349–359, April 2009.

Mohamed L Seghier and Cathy J Price. Visualising inter-subject variability in fMRI using threshold-weighted overlap maps. *Sci. Rep.*, 6:20170, February 2016.

Mohamed L Seghier and Cathy J Price. Interpreting and utilising intersubject variability in brain function. *Trends Cogn. Sci.*, March 2018.

Mohamed L Seghier, Karl J Friston, and Cathy J Price. Detecting subject-specific activations using fuzzy clustering. *Neuroimage*, 36(3):594–605, July 2007.

Mohamed L Seghier, François Lazeyras, Alan J Pegna, Jean-Marie Annoni, and Asaid Khateb. Group analysis and the subject factor in functional magnetic resonance imaging: Analysis of fifty right-handed healthy subjects in a semantic language task. *Hum. Brain Mapp.*, 29(4):461–477, 2008.

Terrence J Sejnowski, Patricia S Churchland, and J Anthony Movshon. Putting big data to good use in neuroscience. *Nat. Neurosci.*, 17(11):1440–1441, November 2014.

Robert Simpson, Kevin R Page, and David De Roure. Zooniverse: observing the world's largest citizen science platform. In *Proceedings of the 23rd International Conference on World Wide Web*, pages 1049–1054. ACM, April 2014.

J G Sled, A P Zijdenbos, and A C Evans. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Trans. Med. Imaging*, 17(1):87–97, February 1998.

Stephen M Smith, Thomas E Nichols, Diego Vidaurre, Anderson M Winkler, Timothy E J Behrens, Matthew F Glasser, Kamil Ugurbil, Deanna M Barch, David C Van Essen, and Karla L Miller. A positive-negative mode of population covariation links brain connectivity, demographics and behavior. *Nat. Neurosci.*, 18(11):1565–1567, November 2015.

José M Soares, Ricardo Magalhães, Pedro S Moreira, Alexandre Sousa, Edward Ganz, Adriana Sampaio, Victor Alves, Paulo Marques, and Nuno Sousa. A hitchhiker's guide to functional magnetic resonance imaging. *Front. Neurosci.*, 10:515, November 2016.

Sebastian M Sodini, Kathryn E Kemper, Naomi R Wray, and Maciej Trzaskowski. Comparison of genotypic and phenotypic correlations: Cheverud's conjecture in humans. *Genetics*, 209(3):941–948, July 2018.

T A Sørensen, T Sørensen, T Sorensen, T J Sørensen, T J Sørensen, T Biering-Sørensen, and L Sorensen. A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on danish commons. January 1948.

Karin Stromswold. The heritability of language: A review and metaanalysis of twin, adoption, and linkage studies. *Language*, 77(4):647–723, 2001.

Cathie Sudlow, John Gallacher, Naomi Allen, Valerie Beral, Paul Burton, John Danesh, Paul Downey, Paul Elliott, Jane Green, Martin Landray, Bette Liu, Paul Matthews, Giok Ong, Jill Pell, Alan Silman, Alan Young, Tim Sprosen, Tim Peakman, and Rory Collins. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.*, 12(3):e1001779, March 2015.

Susan M Sunkin, Lydia Ng, Chris Lau, Tim Dolbeare, Terri L Gilbert, Carol L Thompson, Michael Hawrylycz, and Chinh Dang. Allen brain atlas: an integrated spatio-temporal portal for exploring the central nervous system. *Nucleic Acids Res.*, 41(Database issue): D996–D1008, January 2013.

S Urchs, J Armoza, Y Benhajali, J St-Aubin, and others. MIST: A multi-resolution parcellation of functional brain networks. *MNI Open*, 2017.

Sebastian Urchs, Jonathan Armoza, and Pierre Bellec. dashQC: An interactive quality control dashboard — dashQC_fmri alpha documentation, September 2018.

Sebastian G Urchs, Angela Tam, Pierre Orban, Clara Moreau, Yassine Benhajali, Hien Duy Nguyen, Alan C Evans, and Pierre Bellec. Subtypes of functional connectivity associate robustly with ASD diagnosis. April 2020.

Mara van der Meulen, Nikolaus Steinbeis, Michelle Achterberg, Marinus H van IJzendoorn, and Eveline A Crone. Heritability of neural reactions to social exclusion and prosocial compensation in middle childhood. *Dev. Cogn. Neurosci.*, 34:42–52, November 2018.

David C Van Essen, Stephen M Smith, Deanna M Barch, Timothy E J Behrens, Essa Yacoub, Kamil Ugurbil, and WU-Minn HCP Consortium. The WU-Minn human connectome project: an overview. *Neuroimage*, 80:62–79, May 2013.

John Darrell Van Horn, Scott T Grafton, and Michael B Miller. Individual variability in brain activity: A nuisance or an opportunity? *Brain Imaging Behav.*, 2(4):327–334, December 2008.

Wouter van Rheenen, Wouter J Peyrot, Andrew J Schork, S Hong Lee, and Naomi R Wray. Genetic correlations of polygenic disease traits: from theory to practice. *Nat. Rev. Genet.*, 20(10):567–581, October 2019.

Jakob von Uexküll. A foray into the worlds of animals and humans: with a theory of meaning, trans. *J. O'Neil (Minneapolis, MN: University of Minnesota)*, 2010.

Scott I Vrieze, William G Iacono, and Matt McGue. Confluence of genes, environment, development, and behavior in a post Genome-Wide association study world. *Dev. Psychopathol.*, 24(4):1195–1214, November 2012.

Günter P Wagner and Jianzhi Zhang. The pleiotropic structure of the genotype–phenotype map: the evolvability of complex organisms. *Nat. Rev. Genet.*, 12(3):204–213, March 2011.

Rebecca E Watson, John M Desesso, Mark E Hurtt, and Gregg D Cappon. Postnatal growth and morphological development of the brain: a species comparison. *Birth Defects Res. B Dev. Reprod. Toxicol.*, 77(5):471–484, October 2006.

Wellcome Centre for Human Neuroimaging. SPM12 - statistical parametric mapping. `https://www.fil.ion.ucl.ac.uk/spm/software/spm12/`. Accessed: 2019-4-19.

Tonya White, Philip R Jansen, Ryan L Muetzel, Gustavo Sudre, Hanan El Marroun, Henning Tiemeier, Anqi Qiu, Philip Shaw, Andrew M Michael, and Frank C Verhulst. Automated quality assessment of structural magnetic resonance images in children: Comparison with visual inspection and surface-based reconstruction. *Hum. Brain Mapp.*, 39(3):1218–1231, March 2018.

Susan Whitfield-Gabrieli and Alfonso Nieto-Castanon. Conn: a functional connectivity toolbox for correlated and anticorrelated brain networks. *Brain Connect.*, 2(3):125–141, July

2012.

K J Worsley. An overview and some new developments in the statistical analysis of PET and fMRI data. *Hum. Brain Mapp.*, 5(4):254–258, 1997.

K J Worsley, C Liao, M Grabove, V Petre, B Ha, and A C Evans. A general statistical analysis for fMRI data, 2003.

Zachary A Yaple, W Dale Stevens, and Marie Arsalidou. Meta-analyses of the n-back working memory task: fMRI evidence of age-related changes in prefrontal cortex involvement across the adult lifespan. *Neuroimage*, 196:16–31, August 2019.

Benhajali Yassine and Bellec Pierre. Quality control and assessment of the NIAK functional MRI preprocessing pipeline. *figshare*, November 2016.

Xiuming Zhang, Elizabeth C Mormino, Nanbo Sun, Reisa A Sperling, Mert R Sabuncu, B T Thomas Yeo, Michael W Weiner, Paul Aisen, Michael Weiner, Ronald Petersen, and Others. Bayesian model reveals latent atrophy factors with dissociable cognitive trajectories in alzheimer's disease. *Proceedings of the National Academy of Sciences*, 113 (42):E6535–E6544, 2016.