Université de Montréal

Dealing with heterogeneity in the prediction of clinical diagnosis

par
Christian Langlois Dansereau

Département d'informatique et de recherche opérationnelle
Faculté des arts et des sciences

Thèse présentée à la Faculté des études supérieures
en vue de l'obtention du grade de Philosophiæ Doctor (Ph.D.)
en informatique

Août, 2017

RÉSUMÉ

Le diagnostic assisté par ordinateur est un domaine de recherche en émergence et se situe à l'intersection de l'imagerie médicale et de l'apprentissage machine. Les données médicales sont de nature très hétérogène et nécessitent une attention particulière lorsque l'on veut entraîner des modèles de prédiction. Dans cette thèse, j'ai exploré deux sources d'hétérogénéité, soit l'agrégation multisites et l'hétérogénéité des étiquettes cliniques dans le contexte de l'imagerie par résonance magnétique (IRM) pour le diagnostic de la maladie d'Alzheimer (MA). La première partie de ce travail consiste en une introduction générale sur la MA, l'IRM et les défis de l'apprentissage machine en imagerie médicale. Dans la deuxième partie de ce travail, je présente les trois articles composant la thèse. Enfin, la troisième partie porte sur une discussion des contributions et perspectives futures de ce travail de recherche. Le premier article de cette thèse montre que l'agrégation des données sur plusieurs sites d'acquisition entraîne une certaine perte, comparativement à l'analyse sur un seul site, qui tend à diminuer plus la taille de l'échantillon augmente. Le deuxième article de cette thèse examine la généralisabilité des modèles de prédiction à l'aide de divers schémas de validation croisée. Les résultats montrent que la formation et les essais sur le même ensemble de sites surestiment la précision du modèle, comparativement aux essais sur des nouveaux sites. J'ai également montré que l'entraînement sur un grand nombre de sites améliore la précision sur des nouveaux sites. Le troisième et dernier article porte sur l'hétérogénéité des étiquettes cliniques et propose un nouveau cadre dans lequel il est possible d'identifier un sous-groupe d'individus qui partagent une signature homogène hautement prédictive de la démence liée à la MA. Cette signature se retrouve également chez les patients présentant des symptômes modérés. Les résultats montrent que 90% des sujets portant la signature ont progressé vers la démence en trois ans. Les travaux de cette thèse apportent ainsi de nouvelles contributions à la manière dont nous approchons l'hétérogénéité en diagnostic médical et proposent des pistes de solution pour tirer profit de cette hétérogénéité.

Mots clés Hétérogénéité, Maladie d'Alzeimer, Apprentissage machine, Multisite, Biomarqueur.

ABSTRACT

Computer assisted diagnosis has emerged as a popular area of research at the intersection of medical imaging and machine learning. Medical data are very heterogeneous in nature and therefore require careful attention when one wants to train prediction models. In this thesis, I explored two sources of heterogeneity, multisite aggregation and clinical label heterogeneity, in an application of magnetic resonance imaging to the diagnosis of Alzheimer's disease. In the process, I learned about the feasibility of multisite data aggregation and how to leverage that heterogeneity in order to improve generalizability of prediction models. Part one of the document is a general context introduction to Alzheimer's disease, magnetic resonance imaging, and machine learning challenges in medical imaging. In part two, I present my research through three articles (two published and one in preparation). Finally, part three provides a discussion of my contributions and hints to possible future developments. The first article shows that data aggregation across multiple acquisition sites incurs some loss, compared to single site analysis, that tends to diminish as the sample size increase. These results were obtained through semi-synthetic Monte-Carlo simulations based on real data. The second article investigates the generalizability of prediction models with various cross-validation schemes. I showed that training and testing on the same batch of sites over-estimates the accuracy of the model, compared to testing on unseen sites. However, I also showed that training on a large number of sites improves the accuracy on unseen sites. The third article, on clinical label heterogeneity, proposes a new framework where we can identify a subgroup of individuals that share a homogeneous signature highly predictive of AD dementia. That signature could also be found in patients with mild symptoms, 90% of whom progressed to dementia within three years. The thesis thus makes new contributions to dealing with heterogeneity in medical diagnostic applications and proposes ways to leverage that heterogeneity to our benefit.

Keywords: Heterogeneity, Alzeimer's disease, machine-learning, multisite, biomarker.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

AD	Alzheimer's disease
BOLD	Blood oxygen-level dependent
CN	Cognitively normal
CSF	Cerebrospinal fluid
CV	Cross-validation
fMRI	Functional magnetic resonance imaging
GSC	Global signal correction
HPC	Highly predictable cases
IRM	Imagerie par résonance magnétique
MA	Maladie d'Alzheimer
MCI	Mild cognitive impairment
MR	Magnetic resonance
MRI	Magnetic resonance imaging
NIAK	Neuroimaging analysis kit
PET	Positron-emission tomography
RS	Resting-state
sMRI	Structural magnetic resonance imaging
SNR	Signal to noise ratio

To my parents, for their unwavering support...

ACKNOWLEDGMENTS

My first contact with medical imaging research was 8 years ago at McGill University. I was captivated by the idea of using my technical knowledge to an important and stimulating field that is medicine. I was driven by the idea of helping patients and medical professionals to better understand and diagnose diseases at the individual level. During my studies as a Master student, I had the chance to collaborate with Dr. Pierre Bellec and he eventually convinced me to start a Ph.D. with him. Doing my Ph.D. was probably the best move I did... after asking my fiancé to marry me of course. Thank you so much, Pierre, for your support, guidance, great conversations and for giving me your love of research, I will be forever grateful. I would also like to thank my colleague's from the lab past and present: Aman, Sebastian, Perrine, Yassine, Amal, PO, Jacke, Clara and Hien with whom I got a lot of great exchanges and fun. I would especially like to thank Pierre Orban and Angela for their great advice and support, your friendship and support definitely made a difference. Thanks to the boys of the computer science department: Maor, Thomas, Francis, Cesar, Alexandre, Ishmael, I had extremely stimulating exchanges and ideas with you and I look forward to more of those in the future! Thanks also to my friends outside academia who kept me in a good mental state all those years and were so supportive.

Particular thanks to my parents and my brother Samuel for their great support throughout these years and through my academic endeavor. It has not always been easy but I made it through, thank you. Dad, your passion for science and research was probably the seed of my interest in that field. My in-laws: Isabel, Alain, and Guido who always believed in me and encouraged me to pursue my dreams.

Finally, I would like to give a very special thanks to the love of my life Bianca. Probably my greatest supporter and the one who always had faith in me and push me to surpass myself. Without you, I probably would have starved or eaten a very basic diet of ramens! You did all this even if you had to do your very own Ph.D. at the same time. I admire you...

CHAPTER 1

INTRODUCTION

1.1 General context

Machine learning is on a course to change the way clinical diagnoses are established and delivered. Supervised learning has historically needed large datasets to be able to perform well and unfortunately, this is a scarce resource in medical imaging. One solution to increase the sample size is to aggregate data from heterogeneous sources, with the downside of adding more variance in the dataset. Another source of variance that can impact the performance of an inference model is the imperfect knowledge of clinical diagnoses, reflected in the labels used for training and evaluating our models. Clinical diagnoses are often incorrect, incomplete or not specific enough to the variants that exist in the pathophysiology within a given disorder. Heterogeneous data sources and heterogeneous clinical labels are two issues particularly prevalent in the diagnosis and prognosis of Alzheimer's disease (AD) using magnetic resonance imaging (MRI), which is the main application of my doctoral work.

The number of Canadians suffering from AD is rapidly increasing, with tremendous social and economic impact. Despite the emergence of promising drugs, the recent clinical trials with demented patients have failed. Dementia comes very late in the development of the disease, at a stage where the degeneration of neural tissues has likely gone beyond repair. In order to be efficient, therapies should be initiated in the decades predating dementia, in a preclinical stage where patients experience no or very mild symptoms (see chapter 1.2). There is, unfortunately, no accurate biomarker(s) that can predict AD in this preclinical stage, and that could help identify the individuals that would progress to dementia and benefit from such interventions. It would also be useful to identify presymptomatic markers of the disease in order to understand the underlying mechanism of the pathology. Promising early AD biomarkers can be captured using MRI, which is a broadly available and noninvasive technique. Two separate modalities have been shown

to be of great interest in the investigation of the disease progression, namely structural MRI - that can give information on brain atrophy patterns - and functional MRI - that investigates functional interactions between various brain structures - (see chapter 1.3). Generation of biomarkers require a complex process of data preparation: preprocessing (denoising and spatial alignment) (see chapter 1.4) and features extraction. A standard way to extract meaningful information from the rich 4D images provided by fMRI is to use resting-state connectivity measures and is detailed in chapter 1.5. Different practices like scientific consortia, data sharing and open clinical trials have emerged, and all deliver large public and multisite datasets which can be used to discover new biomarkers (see chapter 1.6). Unfortunately, the gain in sample size due to data aggregation across sites comes at the price of increased heterogeneity (see chapter 1.7.3.1 and 1.7.3.2) and may impact the discriminative properties of our markers.

In addition, heterogeneity also exists in the clinical labels (Drysdale et al. 2017) (see chapter 1.7.3.3). The last point could drastically affect the ability of a prediction model to converge to a solution that will effectively predict clinical labels. Finally I will outline the objectives and contributions of my Ph.D. thesis to deal with technical and clinical sources of heterogeneity in section 1.8.

1.2 Alzheimer's disease

Alzheimer's disease (AD) is a major neurodegenerative disorder characterized by the accumulation of beta amyloid plaques and tau neurofibrillary tangles in the brain. AD gradually destroys a patient's memory and ability to reason, make judgments, communicate and carry out daily activities (Jeong 2004). With the aging of the population worldwide, this disorder has attracted much attention. Evidence from elderly individuals suggests that the pathophysiological process of AD begins years, if not decades, before the diagnosis of clinical dementia (Morris 2005). The clinical disease stages of AD are divided into three phases described by Jack and colleagues Jack et al. (2010).

First is a pre-symptomatic phase in which individuals are cognitively normal but some have pathological changes in AD. Second is a prodromal phase of AD, commonly

referred to as mild cognitive impairment (MCI) (Petersen 2004), which is characterized by the onset of the earliest cognitive symptoms (typically deficits in episodic memory) that do not meet the criteria for dementia. The severity of cognitive impairment in the MCI phase of AD varies from an early manifestation of memory dysfunction to more widespread dysfunction in other cognitive domains. The final phase in the evolution of AD is dementia, defined as multi-domain impairments that are severe enough to result in loss of function (Jack et al. 2010).

The use of a biomarker for the early diagnosis of pathologies has a long history, with many studies showing the feasibility of using an AD biomarker to predict conversion from MCI to AD. These studies show that individuals in the course of developing AD can be identified earlier in the course of the disease by using the MCI stage with the addition of imaging and cerebrospinal fluid (CSF) biomarkers to enhance diagnostic specificity (Chetelat et al. 2003, Jack et al. 1999, Mattsson et al. 2009, Yuan et al. 2009). It could be possible to diagnose AD after the exclusion of other forms of dementia, although a formal diagnosis can currently only be made after a post-mortem evaluation of the brain tissue (McKhann et al. 1984). This is one of the reasons why MRI based analysis and diagnostic tools are currently undergoing intense study in clinical neuroscience research. The early prediction of disease onset is also needed for clinical trials investigating disease-modifying therapies, since treatment of patients with no or mild symptoms are more likely to have a positive outcome, compared to demented subjects who may have such extensive damage that it may be too late to modify the trajectory of the disease.

The current dominant hypothesis in the field for the chain of events in AD pathophysiology is the β-amyloid (Aβ)-cascade. It suggests that interstitial Aβ proteins exert a toxic effect on surrounding neurons and synapses by forming plaques, thereby disturbing their function (Hardy and Selkoe 2002, Shankar et al. 2008). Moreover, a recent research study suggests that, prior to neuronal death resulting in brain atrophy, disruption of functional connectivity may arise in response to an unknown systemic problem and represent an early outcome of Aβ protein plaque formation in AD (Sheline and Raichle 2013). Atrophy is the result of neuronal death and is measured in vivo using structural

MRI measuring the thickness of the gray matter of the cortex (also called cortical thickness) or the gray matter volume in various parcels of the brain. Already in the stage preceding aggregation of Aβ fragments into amyloid plaques, there is a dysfunction of synaptic transmission in many brain regions due to dimers and monomers from the Aβ cascade (D'Amelio and Rossini 2012). As illustrated in Sperling et al. (2011) a viable hypothesis is that functional changes precede the structural changes as well as clinical symptoms and are believed to start in the preclinical phase of the disease. A multimodal combination of structural and functional information may, therefore, lead to accurate predictions of individual clinical trajectories.

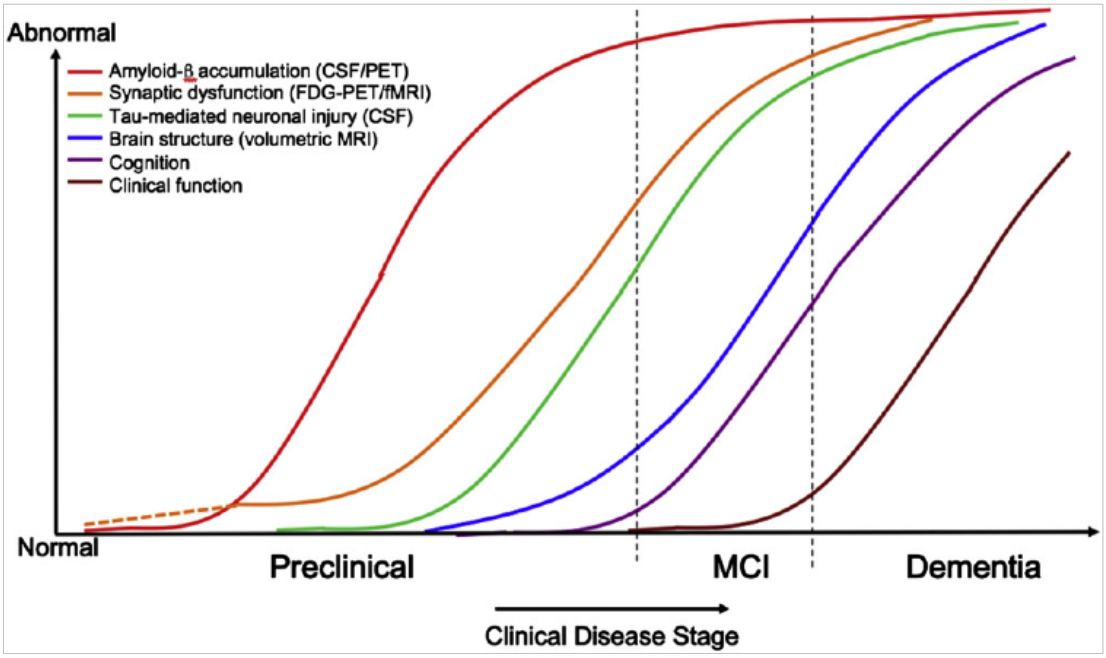

Figure 1.1: Hypothetical model of dynamic biomarkers of the AD expanded to explain the preclinical phase: Aβ as identified by cerebrospinal fluid Aβ42 assays or PET amyloid imaging. Synaptic dysfunction evidenced by fluorodeoxyglucose (F18) positron emission tomography (FDG-PET) or functional magnetic resonance imaging (fMRI), with a dashed line to indicate that synaptic dysfunction may be detectable in carriers of the ε4 alleles of the apolipoprotein E gene before detectable Aβ deposition. Neuronal injury is evidenced by cerebrospinal fluid tau or phospho-tau, and brain structure is evidenced by structural magnetic resonance imaging. Biomarkers change from normal to maximally abnormal (y-axis) as a function of disease stage (x-axis). The temporal trajectory of two key indicators used to stage the disease clinically, cognitive and behavioral measures, and clinical function is also illustrated. Figure from Sperling et al. (2011).

1.3 Overview of magnetic resonance imaging

1.3.1 Overview of structural magnetic resonance imaging

Structural magnetic resonance imaging (sMRI) also called anatomical MRI is an imaging technique that provides static anatomical information. Some atomic nuclei are

able to absorb and emit radio frequency energy when placed in an external magnetic field. In clinical and research MRI, hydrogen atoms are most often used to generate a detectable radio-frequency signal that is received by antennas in close proximity to the head. Hydrogen atoms exist naturally in mammals and in abundance, particularly in water and fat. For this reason, most of the structural MRI sequences essentially map the location of water and fat in the body. Pulses of radio waves excite the nuclear spin of the hydrogen atoms to determine the hydrogen concentration (a proxy for water concentration), and magnetic field gradients are used to localize the signal in space by encoding the radio frequency in space. By varying the parameters of the pulse sequence, different contrasts may be generated between tissues based on the relaxation properties of the hydrogen atoms. The main two contrasts used are $T1-$ and $T2-$ weighted imaging. This imaging modality is widely used in hospitals and clinics for medical diagnosis, staging of disease using brain atrophy (see Figure 1.2) and follow-up without exposing the body to ionizing radiation.

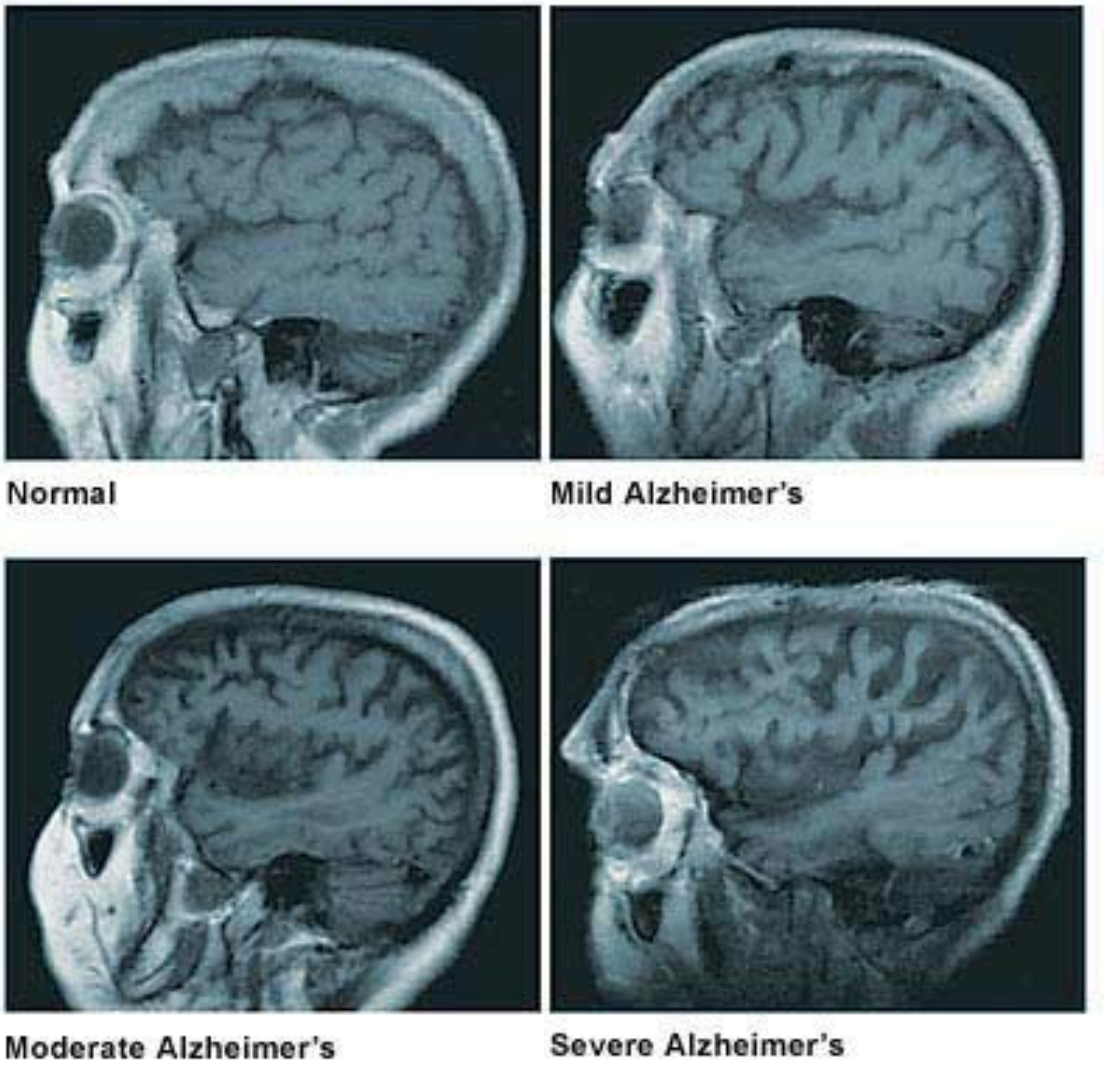

Figure 1.2: The figure for brain atrophy at four stages of AD pathology. Figure from Mayo fondation[1]

1.3.2 Overview of functional magnetic resonance imaging

In functional magnetic resonance imaging (fMRI), the acquisition process is slightly different than for the anatomical MRI acquisition. fMRI uses the principle of the relaxation of hydrogen nuclei, by using specific fMRI sequences of $T2*$ weighted acquisi-

tions that are sensitive to local distortions of the magnetic field. Deoxyhemoglobin, a form of hemoglobin without oxygen, will create such local distortions of the magnetic field, since it is a paramagnetic molecule (positive magnetic susceptibility) (Ogawa et al. 1990). The data acquired using fMRI rely on the hypothesis that areas showing decreased deoxyhemoglobin concentration are due to sustained brain activity. Following neuronal activity, neurons require energy to restore the electrical and ionic concentration balance across the cell membrane. The main mechanism to generate this energy is glucose oxidative metabolism, which requires the delivery of oxygen and glucose by the blood to the site where brain activity takes place (Ogawa et al. 1990). Initially, fMRI was thought to be a good technique to measure the cerebral metabolic rate of oxygen, since the new blood rushing in causes a proportional effect on the venous-end, resulting in a decrease in deoxyhemoglobin concentration, and thus an increase in fMRI signal. The concentration of deoxyhemoglobin actually depends mainly on three factors or phenomena: the metabolic rate of oxygen consumption, cerebral blood volume and cerebral blood flow (Hoge et al. 1999). As a result, the fMRI signal is the outcome of competing effects following neuronal activity.

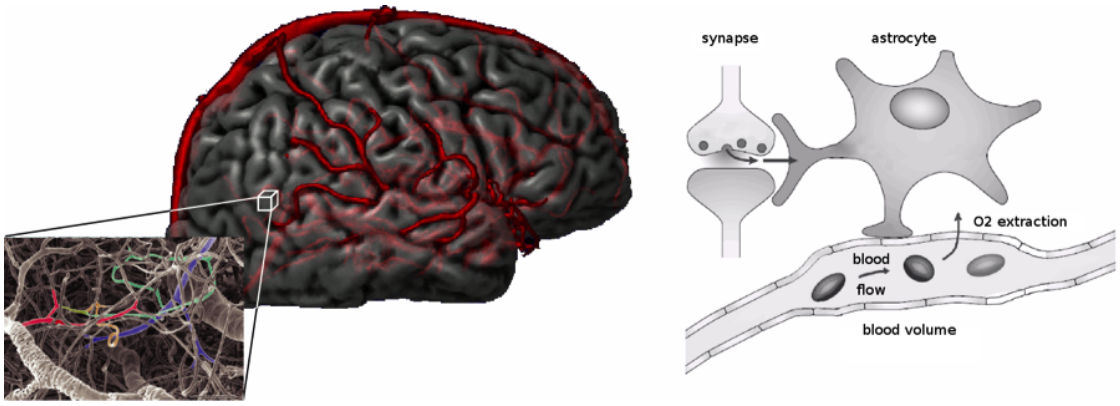

Figure 1.3: Representation of the brain and its vasculature (on the left) and a schematic view of the interaction between the effect of neuronal activity on local changes in blood oxygenation signal (BOLD) (on the right) (adapted from Heeger and Ress (2002)).

1.4 Preprocessing

Normalization of the data is crucial to obtain a consistent and accurate classifier (Kotsiantis 2007). Therefore particular attention is placed on the correction and normalization procedure applied to the rs-fMRI data used in this study. A series of standard preprocessing steps is usually applied in an attempt to correct for various artifacts that would perturb the subsequent analysis and to align the brains of different individuals. The BOLD effect associated with neuronal activity generally results in a relatively small fluctuation of the MR signal. Many factors can influence this signal. Among them, the physiological activity associated mainly with respiration, cardiac pulsations, and patient's motion are major contributors to the noise and are spatially spread everywhere within the brain volume. These sources of noise result in large correlations between BOLD signals of distant voxels. Another factor is the fact that we need a form of spatial normalization of the individual brains in order to perform analysis across subjects (due to anatomical variance among subjects). This spatial normalization (coregistration of the individual brains with a reference template) is necessary but can potentially be another source of confound.

Therefore, preprocessing methods were designed in an attempt to remove specifically the so-called structured noise and motion artifacts from the raw fMRI data. A schematic representation of the preprocessing pipeline can be seen in Figure 1.4.

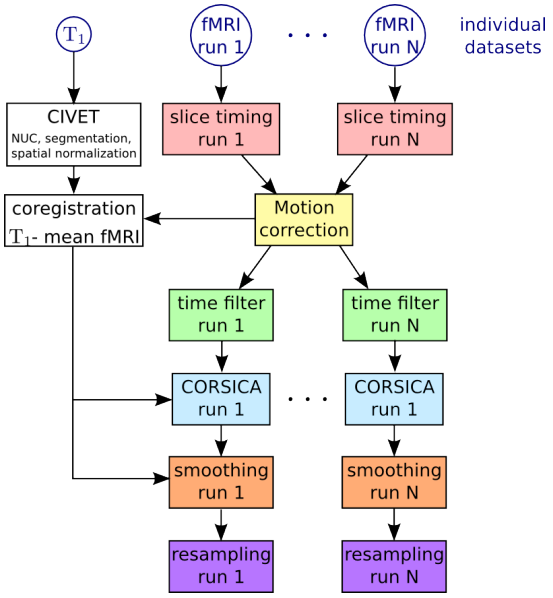

Figure 1.4: Schematic of the preprocessing pipeline including spatial and functional normalization (NIAK preprocessing pipeline [2]).

The basic steps are as follows: (1) correction for slice timing differences due to delay in acquisition sampling; (2) rigid-body motion estimation for within and between runs. Motion correction operates by selecting one functional volume as a reference to align all other functional volumes. Most head motion algorithms describe head movement by 6 parameters, three translation parameters (displacement) and three rotation parameters and are sufficient to characterize the motion of rigid bodies (see Figure 1.5); (3) Coregistration of the functional data in a reference space; (4) resampling of the functional data in the stereotaxic space (references brain used as a common space between subjects); (5) regression of confounds in order to remove spatially structured noise from the fMRI time-series. The confounds are the slow time drift, the high-frequency noise signal, motion parameters, the average signal white matter as well as the average signal of the ventricles (containing cerebrospinal fluid CSF a frequent source of noise and artifact). Some groups have suggested that these corrections are not sufficient to remove motion artefacts and propose some additional corrective procedure (detailed in Chapter 2); and (6) the spatial smoothing is usually applied using a Gaussian blurring kernel to

improve signal to noise ratio (SNR), improve validity of the statistical tests by making the error distribution more normal and finally reduce anatomical and functional variations between subjects (Mikl et al. 2008, Worsley and Friston 1995).

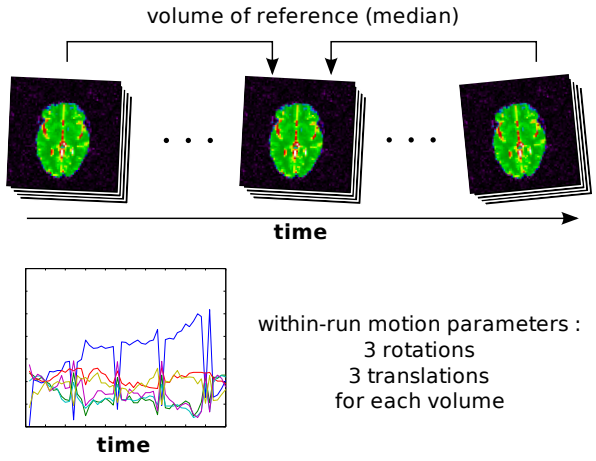

Figure 1.5: Motion estimation based on rigid-body motion estimation of the functional volumes, the procedure provides 6 motion parameters for each volume (3 translation and 3 rotation) Schematic of the preprocessing pipeline including spatial and functional normalization (from NIAK preprocessing pipeline [3]).

1.5 Resting-state connectivity

Resting-state (RS) functional connectivity captures the spatial coherence of slow fluctuations in hemodynamic temporal activity, without performing a prescribed task, as opposed to the task-based acquisition where the subject has to perform a specific task. In resting-state acquisition, the subject is instructed to rest with his eyes open or closed. These temporal fluctuations can be monitored using the signal measured with fMRI. The first study that introduced the concept of resting state functional connectivity was the one of Biswal et al. (1995). They considered the left primary sensorimotor cortex as a seed region for an analysis in a resting-state condition. This analysis consisted in calculating the temporal correlation between the signal of the seed area and the time course of all the voxels of the brain. They found RS correlations between brain regions known to be involved in sensorimotor function. A more recent review done by Fox and Raichle

(2007) illustrated in Figure 1.6 shows the ability to identify the complete sensorimotor network using only the BOLD signal from a small region of that network.

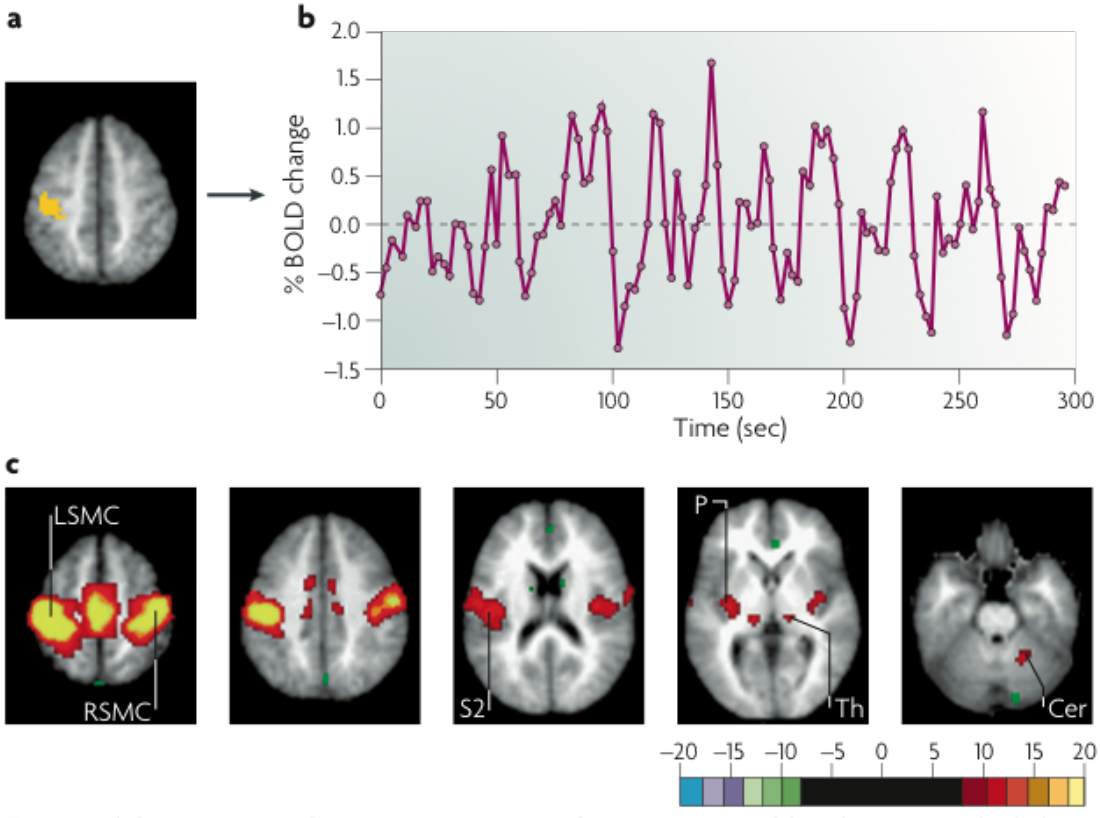

Figure 1.6: Generation of resting-state correlation maps. a) Seed region in the left somatomotor cortex (LSMC) is shown in yellow. b) Time course of spontaneous blood oxygen level dependent (BOLD) activity recorded during resting fixation and extracted from the seed region. c) Statistical z-score map showing voxels that are significantly correlated with the extracted time course. Their significance was assessed using a random effects analysis across a population of ten subjects. In addition to correlations with the right somatomotor cortex (RSMC) and medial motor areas, correlations are observed with the secondary somatosensory association cortex (S2), the posterior nuclei of the thalamus (Th), putamen (P) and cerebellum (Cer) (Fox and Raichle 2007).

These early results from Biswal et al. suggest that it is possible to identify the functional organization of different structures without doing any specific task, just by looking at spontaneous fluctuations in brain activity. Several studies have demonstrated that pat-

terns extracted from temporal correlations of RS signals within the brain volume are organized in space and have a good reproducibility from subject to subject (Damoiseaux et al. 2006, Dansereau et al. 2014). Each network is a combination of multiple brain regions or units, not necessarily spatially close to each other, which share similar low frequency fluctuations of the BOLD signal. This information is usually represented as a functional connectivity matrix where one column of the matrix represents the connectivity of a region or network with the rest of the brain called a functional connectivity map (see Figure 1.7). These networks show the functional organization of various brain regions (see Figure 1.8 for a list of common RS networks).

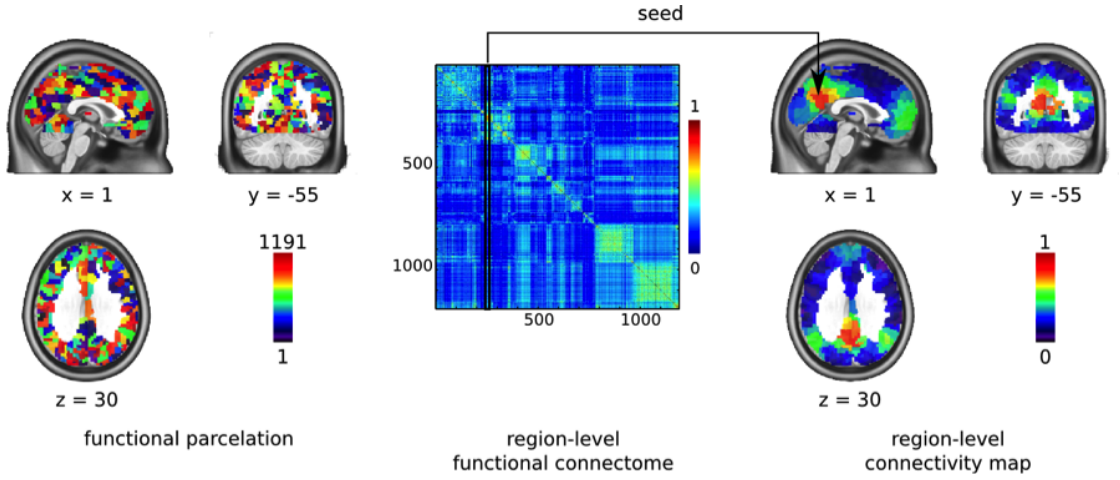

Figure 1.7: Functional connectome: on the left a representation of a functional parcellation, in the middle a region-level functional connectome representing the connectivity between each pair of region, and on the right the connectivity map based on a region of interest extracted from the functional connectome.

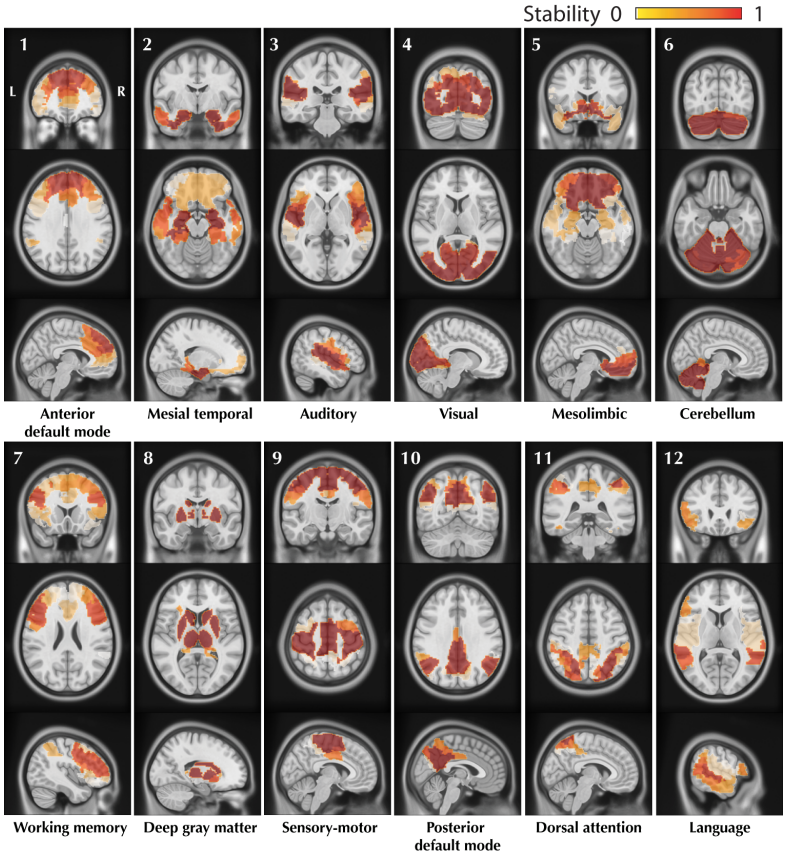

Figure 1.8: The figure shows 12 RS networks identified using a BASC (Bootstrap Analysis of Stable Cluster (Bellec et al. 2010b)) group level analysis of 25 healthy control subjects. BASC is a clustering based method using evidence accumulation for the identification of stable clusters. For each network: 3 slices (coronal, axial, sagittal) are shown superimposed on an anatomical MRI template (MNI152). Labelling of each network was done visually based on previously reported intrinsic networks in the literature. The figure shows networks typically reported in the litterature: Default Mode Network (#1,#10), Auditory (#3), Visual (#4), Sensory-Motor (#9), Attention (#7,#11) and Language(#12). BASC also identified 4 other networks, less often reported, but characterized by high statistical stability: Mesio-Temporal (#2), Mesolimbic (#5), Cerebellum (#6) and Deep Gray Matter (#8) (Dansereau et al. 2014).

Interestingly, RS fMRI signals have also been used for the diagnosis of some neurological and psychiatric disorders, such as multiple sclerosis (Lowe et al. 2002), epilepsy (Waites et al. 2006), schizophrenia (Liang et al. 2006, Salvador et al. 2007, Zhou et al. 2007; 2008), attention deficit hyperactivity disorder (Tian et al. 2007, Zang et al. 2007),

blindness (Liu et al. 2007, Yu et al. 2007), major depression (Anand et al. 2005, Greicius et al. 2007) and acute brainstem ischemia (Salvador et al. 2005). We believe that resting-state fMRI will be an increasingly important modality for exploring the functional abnormalities of patients with AD (Buckley et al. 2017) since we would like to identify signs of the pathology prior to major atrophy or cognitive decline. Therefore functional alterations and/or compensation are believed to occur earlier in the disease progression justifying the use of this modality for early detection of AD pathology.

1.6 Multisite

Quality and quantity of data are usually the main factors that influence the ability of a model to do good inferences. The quantity of data can be expanded through data aggregation, hence why an increasing number of large publicly available cohorts of subjects has emerged, like the 1000 functional connectome (Biswal et al. 2010), ADNI (Mueller et al. 2005), and ABIDE (Di Martino et al. 2014), among others (see Woo et al. (2017) Table 1 for a more exhaustive list). In a clinical trial, the justification for multisite acquisition is more of a logistical one than a financial reason; they need to recruit a large amount of subjects in a short period of time. In order to achieve this goal, they mandate the recruitment to multiple clinical centers across the globe which accelerates the evaluation time of a drug. Although these centers may have been harmonized by their scanner protocols, scanners will have differences in their software version, the specific add-on to the scanners, and, most importantly, vendors (even field strength may differ in some cases). Unfortunately, between studies, MR acquisition methodologies are among the most commonly cited sources of measurement variation (Friedman et al. 2006). This is why it is important to assess if multisite resting-state connectivity analysis is feasible: Can we combine the data from multiple sources while still detecting effects of interest in the data? What corrective measure on the data should be applied to reduce the bias introduced by multisite data collection? Among the factors of variability across sites, we can list the following 3 categories described in (Yan et al. 2013b):

1. *Acquisition-related variations:*

    (a) *Scanner make and model (Friedman et al. 2006)*

    (b) *Sequence type (spiral vs. echo planar; single-echo vs. multi-echo) (Klarhofer et al. 2002), parallel vs. conventional acquisition (Feinberg et al. 2010) (Lin et al. 2005)*

    (c) *Coil type (surface vs. volume, number of channels, orientation).*

    (d) *Acquisition parameters: repetition time, number of repetitions, flip angle, echo time, and acquisition volume (field of view, voxel size, slice thickness/-gaps, slice prescription) (Friedman and Glover 2006).*

2. *Experimental-related variations:*

    (a) *Participant instructions (Hartstra et al. 2011), eyes-open/eyes-closed (Yan et al. 2009) (Yang et al. 2007), visual displays, experiment duration (Fang et al. 2007) (Van Dijk et al. 2010).*

3. *Environment-related variations:*

    (a) *Sound attenuation measures (Cho et al. 1998) (Elliott et al. 1999).*

    (b) *Attempts to improve participant comfort during scans (e.g., music, videos) (Cullen et al. 2009).*

    (c) *Head-motion restraint techniques (e.g., vacuum pad, foam pad, bite-bar, plaster cast head holder) (Edward et al. 2000) (Menon et al. 1997).*

    (d) *Room temperature and moisture (Vanhoutte et al. 2006).*

In 2009, the 1000 Functional Connectomes Project (FCP) released a sample of 1200+ fMRI scans collected at 33 different institutions, which provided a glimpse of the variability in imaging characteristics employed by the neuroimaging field. The parameters of the imaging acquisition varied across sites, while the majority of subject-related variables are not reported (due in most cases, to the fact that they were not thoroughly

recorded). Despite justifiable skepticism, feasibility analyses demonstrated that meaningful explorations of the aggregated dataset could be performed (Biswal et al. 2010). Although no explicit correction for multisite variability was used, they only used a global signal correction (GSC) to normalize subjects which may introduce anti-correlations in the data (Carbonell et al. 2014, Fox et al. 2009, Murphy et al. 2009, Power et al. 2014, Saad et al. 2012). After accounting for site-related differences, the analysis showed brain-behavior relationships with variables such as age, gender, and diagnostic label, and confirmed a variety of prior hypotheses (Biswal et al. 2010, Fair et al. 2012, Tomasi and Volkow 2010, Zuo et al. 2012). While encouraging, it remains a source of concern if unharmonized datasets can be aggregated together since many uncontrolled and unknown factors in the 1000 FCP, may be adding variance unrelated to simple site effects as highlighted by Yan et al. (2013a). Another demonstration of substantial multisite variance is the study reported by Nielsen et al. (2013) where the authors compared a monosite and a multisite dataset of subjects with autism and concluded that the multisite classification accuracy was much lower for multisite than monosite (Nielsen et al. 2013).

1.7 Prediction of clinical diagnosis using medical images

1.7.1 Prediction in the context of AD

In the past few years, several major studies have been initiated that have aimed to predict who will develop AD dementia at the prodromal or even asymptomatic stages, with the ultimate goal of providing a platform for interventions with disease-modifying therapies. Many of these studies were designed to evaluate the role of neuroimaging and clinical biomarkers in assessing and predicting progression in individuals without cognitive impairment and in individuals with MCI.

A recent body of literature has focussed on the classification of subjects at various stages of AD and progression from a prodromal stage to AD dementia using one of the various brain imaging modalities available, including positron emission tomography (PET) imaging (Cabral et al. 2015, Mathotaarachchi et al. 2017), structural MRI (Adaszewski et al. 2013, Eskildsen et al. 2013, Liu et al. 2015; 2013, Misra et al. 2009,

Salvatore et al. 2015), and functional MRI Challis et al. (2015), Chen et al. (2011), Jie et al. (2014), Khazaee et al. (2015), see (Rathore et al. 2017) for a complete review. More recently some groups used multimodal data from the ADNI dataset to improve prediction accuracy and reported classification performance of the order of 95% accuracy to classify patients with AD dementia vs. CN (Xu et al. 2015, Zhu et al. 2014, Zu et al. 2016) and 80% accuracy to identify patients with MCI who will progress to AD dementia (Cheng et al. 2015a;b, Korolev et al. 2016, Moradi et al. 2015).

1.7.2 Cross-validation

As machine learning algorithms are increasingly used to support clinical decision making, it is important to reliably quantify how a prediction model will generalize to an independent dataset or site. Cross-validation (CV) is the standard approach for evaluating the accuracy of such algorithms. Validation is the task of training the data on a subset of the data and evaluating its performance on the hold-out portion that has never been seen by the model. Cross-validation, on the other hand, is the repeated measure of the validation with non-overlapping subsamples. Multiple CV methods exist and, depending on the end goal, some may be better than others to obtain an unbiased estimate of the model accuracy as reported by (Saeb et al. 2016) where they compared two popular CV methods: record-wise and subject-wise cross-validation. In their paper, Saeb and colleagues made the case that record-wise CV leads to overestimated accuracy score that does not reflect the true prediction accuracy when evaluated on unseen subjects. As highlighted by Little et al. (2017), Varoquaux (2017) the context and question that we want to answer will determine the optimal CV scheme.

1.7.3 Dealing with heterogeneity

Accounting for heterogeneous sources of variance is important since they may reduce the predictive potential of our model.

1.7.3.1 Feasibility of multisite studies

In order to extract good and reliable biomarkers most machine-learning models require large sample sizes. Neuroimaging analysis is typically being acquired on one single site with the same device. The major limitations of this type of recruitment scheme is the decimation of small samples (small datasets acquired in parallel with different protocols on different scanners that increase the number of studies but not the sample size) and the difficulty to recruit a large amount of participants in the same location in a reasonable time frame as mentioned previously in the multisite section. A solution for the previously mentioned issues is to aggregate data from multiple sites into one large dataset. As referenced earlier, data aggregation poses a number of difficulties due to the added variability incurred by pooling multiple data from multiple sources. The question is: does the tradeoff of having a larger sample supersede the added heterogeneity obtained by aggregating multiple sites into one large dataset?

As indicated earlier, the problem regarding multisite heterogeneity can easily be transposed to a more general problem in machine learning related to the aggregation of various datasets that were not obtained with the same equipment and standards e.g. pictures taken with various digital camera brands of a diverse range of quality and where all picture labels are not uniformly distributed across cameras (Deng et al. 2009). Given sufficient data and a model with enough capacity, this variability can be modeled. We, unfortunately, do not have enough data (the data is very scarce and expensive to acquire) to fully model this variability in medical imaging applications. The lack of ground truth to evaluate the performance of a model in various training configurations and misdiagnosis may affect our ability to evaluate the variance contribution of a multisite acquisition. We, therefore, need to evaluate the detrimental effect of the aggregation of data on our inference models, compared to the standard monosite analysis, using realistic simulations and this is precisely what we propose to do in the context of fMRI data analysis.

1.7.3.2 Generalizability of models: Harnessing heterogeneity

In order to be useful, biomarkers that are identified on one dataset need to be generalizable to other datasets. A standard approach to estimate the performance of a given prediction model on unseen samples is to do cross-validation. This is very well known principle in the machine learning community and a standard step in the evaluation of a prediction model (Friedman et al. 2001). Now let us assume that our data is collected using multiple sources (e.g. recording devices). Multiple scenarios are possible, so we could use only the data coming from one device and do the cross-validation for that dataset to obtain an accuracy score. This accuracy score will reflect the performance of the model for data coming from that specific device but does not give any idea of the performance of the model for data coming from a different device. The same problem can be directly applied in neuroimaging where the dataset is obtained from MRI scanners that may have different properties and site specific characteristics. A lot of the early work on pathology prediction in neuroimaging has reported an accuracy score where the training and test data came from the same scanner (Arbabshirani et al. 2017, Costafreda et al. 2009, Fan et al. 2005, Fu et al. 2008, Hahn et al. 2011, Kawasaki et al. 2007, Lao et al. 2004, Marquand et al. 2008, Mourao-Miranda et al. 2005, Nouretdinov et al. 2011, Rathore et al. 2017). Unfortunately, more recent work has shown that the accuracy drops dramatically when the model is applied to another independent dataset (Abraham et al. 2016, Cheng et al. 2015c, Schilbach et al. 2016, Skåtun et al. 2016, Woo et al. 2017). It is therefore important to evaluate if this behavior was due to a lack of information at training, rendering the model to be unable to generalize well to independent devices and sites. In order to explore this question, it is possible to use various sampling strategies to evaluate if the pooling of multiple sites at training yields a better generalization outcome for unseen data. This is particularly important for clinical use since the predictive model will most likely be used with data obtained from a different site than the site(s) used for training and accuracy evaluation. We, therefore, need a less biased accuracy estimate that will correctly quantify the generalizability of the model in unseen sites. The site heterogeneity needs to be learned while training the model so that the model can become

invariant to it.

1.7.3.3 Clinical label heterogeneity

Heterogeneity can appear in different forms and at different levels, for example, we can have acquisition heterogeneity due to: recruitment bias, the use of different scanners, software to process the data, interindividual biological differences or any sort of acquisition noise present in the data. Those types of heterogeneity, mainly explained in the previous sections (section 1.7.3.2, and 1.7.3.2), reduce the effect size and therefore may need to be modeled as much as possible to reduce their impact. Another source of heterogeneity is the labels heterogeneity. By labels heterogeneity, we mean that they are not precise enough to encompass the underlying variability of the data or that some sub groups are underrepresented.

This other source of heterogeneity is well established in the clinical world but is usually not accounted for until recently in the machine learning world applied to medical problems. In an ideal scenario you would have multiple sub-diseases in a dataset with a large number of examples of each sub category. In that context the algorithm would learn to identify what is common among all of those subjects even if they have drastically different underlying causes. Unfortunately, in most clinical datasets we are far from unlimited data and a subgroup may be under-represented or simply not identified in some cases, crippling the ability of the model to do its job correctly. The fact that the labels are poorly defined renders the task of a perfect prediction virtually impossible since it is ill posed from the start. We basically have imposed overly strict and sometimes subjective categorical labels to disorders like schizophrenia Insel (2010) (that is more seen as a spectrum disorder) or Alzheimer's disease that encompasses multiple sub-forms of the disease (Lam et al. 2013) and/or mixed pathologies.

Since clinical diagnoses are often incorrect, incomplete or not specific enough to the variants that exist in the pathophysiology, it will impair our ability to have true gold standard labels and inevitably the prediction model will be affected by this lack of precision in the clinical labels. For example Beach et al. (2012) have shown that a clinical error in diagnosis for AD dementia exists after post-mortem neuropathological inves-

tigation revealed that only 70.9% to 87.3% (depending on the clinical criteria) of the probable AD subjects were diagnosed correctly. Beach and colleagues also found that a range between 44.3% to 70.8% (depending on the clinical criteria) of the subjects diagnosed as non-AD had, in fact, AD pathology, as defined by post-mortem histopathologic evaluations. Data-driven analysis of sMRI in AD further showed that symptomatic heterogeneity is related to different patterns of atrophy spreading in AD (Dong et al. 2016, Zhang et al. 2016). Recently, Dong et al. (2016) also reported multiple subtypes of dysconnectivity in patients suffering from AD dementia, MCI, and subjective cognitive impairment, using diffusion magnetic resonance imaging, and reported associations between subtypes and the severity of cognitive impairment. These findings highlight the existence of a great heterogeneity in the signature of AD pathology.

Most classification models propose a built-in confidence estimate over their prediction. Unfortunately, it is possible for a model to be very confident about a prediction that is completely wrong (Niculescu-Mizil and Caruana 2005) this can be true with outliers for example or in over-fitting scenarios (Waterhouse et al. 1996). This is mainly due to the way confidence is calculated, namely the distance of that sample to the hyperplane. To deal with outliers a field of statistics called robust statistic has emerged to focus on that particular issue and render the model to be robust to outliers. Outliers, by definition, usually represent a very small fraction of the examples, and it is precisely in those conditions that the robust statistic holds (Black and Rangarajan 1996). In our case, although we have a heterogeneous population, there is no guarantee that a majority of the subjects are homogeneous, rendering this type of solution unhelpful. The parametric estimation of the model confidence probability that was proposed by Platt et al. (1999), Wu et al. (2004) is limited by strong a priori on the data. We would therefore benefit from a non-parametric metric that could compute the likelihood of a subject to be correctly classified and use that to identify a highly predictable subgroup of subjects.

The label heterogeneity could be better modeled with a very large dataset encompassing most of the labels' variability. This would help in part to refine clinical labels based on groups of individuals who share a common phenotypic signature and better model the intersubject variability. Unfortunately, such large datasets do not exist yet but

may arise in the future. In the meantime what can be done with the existing datasets? Would it be possible to extract meaningful information that can be used in a clinical setup even though they encompass heterogeneous clinical labels?

I propose this approach in the context of AD clinical trial enrichment, but the issue raised is general and touches many machine learning applications. Any high-risk problem where some specific action is very impactful or costly, and the lack of action has a small cost, is relevant. One would want to only implement those actions in the cases where the positive outcome is highly probable. For example, when trading stocks, we would like to place an order to buy or sell a stock only when it is highly probable that the stock will rise or fall in the next time point instead of placing a bet at each time point.

The first part of my scientific contributions is related to realistic multisite simulation and generalizability, and I have proposed a domain specific solution to a general problem. For the second part related to the problem of labels heterogeneity, I have proposed a generic solution to a domain specific problem. Since this solution is generic it could be used in a variety of other domains.

1.8 Objectives

The overall objective of this thesis was to explore the impact of heterogeneity in its various forms on imaging analysis and corrective approaches that can be used to reduce its impact. We mainly focus on two aspects of variance, namely the multisite aggregation and the clinical labels heterogeneity.

1.8.1 First paper objectives

The first contribution of this thesis (Dansereau et al. (2017)) addressed the feasibility and impact of multisite fMRI analysis in standard univariate or multivariate machine learning experiments. This question is very important since it is an emerging strategy to increase sample size and is gaining a lot of interest in the neuroimaging community. Since we lack a ground truth where we have the exact same effect and subjects scanned using a monosite and a multisite scenario, the objective was to use realistic simulations

to generate an equivalent dataset.

1.8.2 Second paper objectives

The second paper's objective (Orban et al. (2017a)) was to explore the generalizability of various training schemes (monosite CV, intra-sites CV, inter-sites CV) in the context of a multisite dataset to identify if a bias exists in the reported accuracy performance. We also aimed to determine the most unbiased strategy to estimate the generalizability performance of a model on true unseen data coming from a different site. Our hypothesis was that even though a multisite acquisition may increase the heterogeneity of the dataset, it is useful to test the generalizability of the results across different samples, making it more likely to obtain generalizable features reflecting generic traits of the pathology rather than particularities of a single dataset. To do so we evaluated the effect of intra-site vs. inter-site training on prediction accuracy performance using real data from a clinical population acquired on different sites instead of simulated effects of pathology like we did in my preceding work.

1.8.3 Third paper objectives

The heterogeneity in clinical labels has to our knowledge not been previously addressed even though it is well-known in the clinical community that diseases like AD may encompass multiple sub diseases that are currently not diagnosed or identified as such. The main problems are the comorbid factors and mismatch between pathological and clinical stages that may cause heterogeneity in the clinical labels. We, therefore, proposed in the third paper to design a prediction pipeline for the data-driven identification of a signature of AD that will account for the heterogeneity of labels and improve the prediction accuracy on a subset of the population. We will also evaluated if that signature can be found in a prodromal stage of the disease (MCI) and if it could be a reliable marker of progression to AD dementia.

CHAPTER 2

STATISTICAL POWER AND PREDICTION ACCURACY IN MULTISITE RESTING-STATE FMRI CONNECTIVITY

Published in Neuroimage. 2017[1]

C. Dansereau, Y. Benhajali, C. Risterucci, E. Merlo Pich, P. Orban, D. Arnold, P. Bellec

2.1 Abstract

Connectivity studies using resting-state functional magnetic resonance imaging are increasingly pooling data acquired at multiple sites. While this may allow investigators to speed up recruitment or increase sample size, multisite studies also potentially introduce systematic biases in connectivity measures across sites. In this work, we measure the inter-site effect in connectivity and its impact on our ability to detect individual and group differences. Our study was based on real, as opposed to simulated, multisite fMRI datasets collected in $N = 345$ young, healthy subjects across 8 scanning sites with 3T scanners and heterogeneous scanning protocols, drawn from the 1000 functional connectome project. We first empirically show that typical functional networks were reliably found at the group level in all sites, and that the amplitude of the inter-site effects was small to moderate, with a Cohen's effect size below 0.5 on average across brain connections. We then implemented a series of Monte-Carlo simulations, based on real data, to evaluate the impact of the multisite effects on detection power in statistical tests comparing two groups (with and without the effect) using a general linear model, as well as on the prediction of group labels with a support-vector machine. As a reference, we also implemented the same simulations with fMRI data collected at a single site using an identical sample size. Simulations revealed that using data from heterogeneous sites

[1] http://dx.doi.org/10.1016/j.neuroimage.2017.01.072

only slightly decreased our ability to detect changes compared to a monosite study with the GLM, and had a greater impact on prediction accuracy. However, the deleterious effect of multisite data pooling tended to decrease as the total sample size increased, to a point where differences between monosite and multisite simulations were small with $N = 120$ subjects. Taken together, our results support the feasibility of multisite studies in rs-fMRI provided the sample size is large enough.

Highlights

- Small to moderate systematic site effects in fMRI connectivity.
- Small impact of site effects on the detection of group differences for sample size > 100.
- Linear regression of the sites prior to multivariate prediction do not improve prediction accuracy.

2.2 Introduction

2.2.1 Main objective

Multisite studies are becoming increasingly common in resting-state functional magnetic resonance imaging (rs-fMRI). In particular, some consortia have retrospectively pooled rs-fMRI data from multiple independent studies comparing clinical cohorts with control groups, e.g. normal controls in the 1000 functional connectome project (FCP) (Biswal et al. 2010), children and adolescents suffering from attention deficit hyperactivity disorder from the ADHD200 (Fair et al. 2012, Milham et al. 2012), individuals diagnosed with autism spectrum disorder in ABIDE (Nielsen et al. 2013), individuals suffering from schizophrenia (Cheng et al. 2015c), or elderly subjects suffering from mild cognitive impairment (Tam et al. 2015). The rationale behind such initiatives is to dramatically increase the sample size at the cost of decreased sample homogeneity. The systematic variations of connectivity measures derived using different scanners, called

site effects, may decrease the statistical power of group comparisons, and somewhat mitigate the benefits of having a large sample size (Brown et al. 2011, Jovicich et al. 2016). In this work, our main objective was to quantitatively assess the impact of site effects on group comparisons in rs-fMRI connectivity.

2.2.2 Group comparison in rs-fMRI connectivity

In this work, we focused on the most common measure of individual functional connectivity, which is the Pearson's correlation coefficient between the average rs-fMRI time series of two brain regions. To compare two groups, a general linear model (GLM) is typically used to establish the statistical significance of the difference in average connectivity between the groups. Finally a p-value is generated for each connection to quantify the probability that the difference in average connectivity is significantly different from zero (Worsley and Friston 1995, Yan et al. 2013b). If the estimated p-value is smaller than a prescribed tolerable level of false-positive findings (see for more detail Table 2.I), generally adjusted for the number of tests performed across connections, say $\alpha = 0.001$, then the difference in connectivity is deemed significant.

2.2.3 Statistical power in group comparisons at multiple sites

The statistical power of a group comparison study is the probability of finding a significant difference, when there is indeed a true difference. A careful study design

		Detected value	
		patho	**no patho**
Actual value	**patho**	True Positive	False Negative
	no patho	**False Positive**	True Negative

Table 2.I: Confusion matrix.

involves the selection of a sample size that is large enough to reach a set level of statistical power, e.g. 80%. In the GLM, the statistical power actually depends on a series of parameters (Desmond and Glover 2002, Durnez et al. 2014): (1) the sample size (the larger the better); (2) the absolute size of the group difference (the larger the better), and, (3) the intrinsic variability of measurements (the smaller the better) (4) the rejection threshold α for the null hypothesis.

2.2.4 Sources of variability: factors inherent to the scanning protocol

In a multisite (or multi-protocol) setting, differences in imaging or study parameters may add variance to rs-fMRI measures, e.g. the scanner make and model (Friedman et al. 2006; 2008), repetition time, flip angle, voxel resolution or acquisition volume (Friedman and Glover 2006), experimental design such as eyes-open/eyes-closed (Yan et al. 2009), experiment duration (Van Dijk et al. 2010), and scanning environment such as sound attenuation measures (Elliott et al. 1999), or head-motion restraint techniques (Edward et al. 2000, Van Dijk et al. 2012), amongst others. These parameters can be harmonized to some extent, but differences are unavoidable in large multisite studies. The recent work of Yan et al. (2013b) has indeed demonstrated the presence of significant site effects in rs-fMRI measures in the 1000 FCP. Site effects will increase the variability of measures, and thus decrease statistical power. To the best of our knowledge, it is not yet known how important this decrease in statistical power may be.

2.2.5 Sources of variability: within-subject

The relative importance of site effects in rs-fMRI connectivity depends on the amplitude of the many other sources of variance. First, rs-fMRI connectivity only has moderate-to-good test-retest reliability using standard 10-minute imaging protocols (Shehzad et al. 2009), even when using a single scanner and imaging session. Differences in functional connectivity across subjects are also known to correlate with a myriad of behavioural and demographic subject characteristics (Anand et al. 2007, Kilpatrick et al. 2006, Sheline et al. 2010). Taken together, these sources of variance reflect a fundamen-

tal volatility of human physiological signals.

2.2.6 Sources of variability: factors inherent to the site

In addition to physiology, some imaging artefacts will vary systematically from session to session, even at a single site. For example, intensity non-uniformities across the brain depend on the positioning of subjects (Caramanos et al. 2010). Room temperature has also been shown to impact MRI measures (Vanhoutte et al. 2006). Given the good consistency of key findings in resting-state connectivity across sites, such as the organization of distributed brain networks (Biswal et al. 2010), it is reasonable to hypothesize that site effects will be small compared to the combination of physiological and within-site imaging variance.

2.2.7 Multivariate analysis

Another important consideration regarding the impact of site effects on group comparison in rs-fMRI connectivity is the type of method used to identify differences. The concept of statistical power is very well established in the GLM framework, which tests one brain connection at a time (mass univariate testing). However, multivariate methods that combine several or all connectivity values in a single prediction are also widely used and likely affected by the site effects. A popular multivariate technique in rs-fMRI is support-vector machine (SVM) (Cortes and Vapnik 1995). In this approach, the group sample is split into a training set and a test set. The SVM is trained to predict group labels on the training set, and the accuracy of the prediction is evaluated independently on the test set. The accuracy level of the SVM captures the quality of the prediction of clinical labels from resting-state connectivity, but does not explicitly tell which brain connection is critical for the prediction. The accuracy score can thus be seen as a "separability index" between the individuals of two groups in high dimensional space. Altogether, the objectives and measures of statistical risk for SVM and GLM are quite different. Because SVM has the ability to combine measures across connections, unlike univariate GLM tests, we hypothesized that the GLM and SVM will be impacted differently by

site effects. Even though the accuracy is expected to be lower for the multisite than the monosite configuration, it as been shown that the generalizability of a predictive model to unseen sites is greater for models trained on multisite than monosite datasets as shown by Abraham et al. (2016).

2.2.8 Specific objectives

Our first objective was to characterize, using real data, the amplitude of systematic site effects in rs-fMRI connectivity measures across sites, as a function of within-site variance. We based our evaluation on images generated from independent groups at 8 sites equipped with 3T scanners, in a subset ($N = 345$) of the 1000 FCP. Our second objective was to evaluate the impact of site effects on the detection power of group differences in rs-fMRI connectivity. To answer this question directly, one would need to scan two different cohorts of participants at least twice, once in a multisite setting and once in a monosite setting. Such an experiment may be too costly to implement for addressing a purely technical objective. As a more feasible alternative, we implemented a series of Monte Carlo simulations, adding synthetic “pathological” effects in the 1000 FCP sample. One interesting feature of the "1000 FCP" dataset is the presence of one large site of ~ 200 subjects and 7 small sites of ~ 20 subjects per site. We were therefore able to implement realistic scenarios following either a monosite or a multisite design (with 7 sites), with the same total sample size. Our simulations gave us full control on critical aspects for the detection of group differences, such as the amplitude of the group difference, sample size, and the balancing of groups across sites. We evaluated the ability of detecting group differences both in terms of sensitivity for a GLM and in terms of accuracy for a SVM model.

2.3 Method

2.3.1 Imaging sample characteristics

The full 1000 FCP sample includes 1082 subjects, with images acquired over 33 sites spread across North America, Europe, Australia and China. As the 1000 FCP is a

retrospective study, no effort was made to harmonize population characteristics or imaging acquisition parameters (Biswal et al. 2010). A subset of sites was selected based on the following criteria: (1) 3T scanner field strength, (2) full brain coverage for the rs-fMRI scan, and, (3) a minimum of 15 young or middle aged adult participants, with a mixture of males and females (4) samples drawn from a population with a predominant Caucasian ethnicity. In addition, only young and middle aged participants (18-46 years old) were included in the study, and we further excluded subjects with excessive motion (see next Section). The final sample for our study thus included 345 cognitively normal young adults (150 males, age range: 18-46 years, mean$\pm$std: 23.8 $\pm$5.14) with images acquired across 8 sites located in Germany, the United Kingdom, Australia and the United States of America. The total time of available rs-fMRI data for these subjects ranged between 6 and 7.5 min and only one run was available per subject. See Table 2.II for more details on the demographics and imaging parameters at each site selected in the study. The experimental protocols for all datasets as well as data sharing in the 1000 FCP were approved by the respective ethics committees of each site. This secondary analysis of the 1000 FCP sample was approved by the local ethics committee at CRIUGM, University of Montreal, QC, Canada.

2.3.2 Computational environment

All experiments were performed using the NeuroImaging Analysis Kit, NIAK[2] (Bellec et al. 2011) version 0.12.18, under CentOS version 6.3 with Octave[3] version 3.8.1 and the Minc toolkit[4] version 0.3.18. Analyses were executed in parallel on the "Mammouth" supercomputer[5] , using the pipeline system for Octave and Matlab, PSOM (Bellec et al. 2012) version 1.0.2. The scripts used for processing can be found on Github[6]. Prediction was performed using the LibSVM library (Chang and Lin 2011). Visualiza-

[2] http://simexp.github.io/niak/

[3] http://gnu.octave.org/

[4] http://www.bic.mni.mcgill.ca/ServicesSoftware/ServicesSoftwareMincToolKit

[5] http://www.calculquebec.ca/index.php/en/resources/compute-servers/mammouth-serie-ii

[6] https://github.com/SIMEXP/Projects/tree/master/multisite

Site	Magnet	Scanner	Channels	N	Nfinal	Sex	Age	TR	#Slices	#Frames
Baltimore, USA	3T	Philips Achieva	8	23	21	8M/15F	20-40	2.5	47	123
Berlin, DE	3T	Siemens Tim Trio	12	26	26	13M/13F	23-44	2.3	34	195
Cambridge, USA	3T	Siemens Tim Trio	12	198	195	75M/123F	18-30	3	47	119
Newark, USA	3T	Siemens Allegra	12	19	17	9M/10F	21-39	2	32	135
NewYork_b, USA	3T	Siemens Allegra	1	20	18	8M/12F	18-46	2	33	175
Oxford, UK	3T	Siemens Tim Trio	12	22	20	12M/10F	20-35	2	34	175
Queensland, AU	3T	Bruker	1	19	17	11M/8F	20-34	2.1	36	190
SaintLouis, USA	3T	Siemens Tim Trio	12	31	31	14M/17F	21-29	2.5	32	127

Table 2.II: Sites selected from the 1000 Functional Connectome Project.

tion was implemented using Python 2.7.9 from the Anaconda 2.2.0[7] distribution, along with Matplotlib[8] (Hunter 2007), Seaborn[9] and Nilearn[10] for brain map visualizations.

2.3.3 Preprocessing

Each fMRI dataset was corrected for slice timing; a rigid-body motion was then estimated for each time frame, both within and between runs, as well as between one fMRI run and the T1 scan for each subject (Collins et al. 1994). The T1 scan was itself nonlinearly co-registered to the Montreal Neurological Institute (MNI) ICBM152 stereotaxic symmetric template (Fonov et al. 2011), using the CIVET pipeline (Ad-Dab'bagh et al. 2006a). The rigid-body, fMRI-to-T1 and T1-to-stereotaxic transformations were all combined to re-sample the fMRI in MNI space at a 3 mm isotropic resolution. To minimize artifacts due to excessive motion, all time frames showing a frame displacement, as defined in Power et al. (2012), greater than 0.5 mm were removed and a residual motion estimated after scrubbing. A minimum of 50 unscrubbed volumes per run was required for further analysis (13 subjects were rejected). The following nuisance covariates were regressed out from fMRI time series: slow time drifts (basis of discrete

[7] http://docs.continuum.io/anaconda/index
[8] http://matplotlib.org/
[9] http://stanford.edu/~mwaskom/software/seaborn/index.html
[10] http://nilearn.github.io/

cosines with a 0.01 Hz highpass cut-off), average signals in conservative masks of the white matter and the lateral ventricles (average Pearson correlation across all subjects is 0.242 between gray matter and white matter signals, and 0.031 between gray matter and ventricles signals) 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, Lund et al. 2006). The fMRI volumes were finally spatially smoothed with a 6 mm isotropic Gaussian blurring kernel. A more detailed description of the pipeline can be found on the NIAK website[11] and Github[12].

2.3.4 Inter-site bias in resting-state connectivity

2.3.4.1 Functional connectomes

We compared the functional connectivity measures derived from different sites of the 1000 FCP. A functional brain parcellation with 100 regions was first generated using a bootstrap analysis of stable clusters (Bellec et al. 2010b), on the Cambridge cohort of the 1000 FCP ($N = 195$), as described in Orban et al. (2015). For a given pair of regions, the connectivity measure was defined by the Fisher transformation of the Pearson's correlation coefficient between the average temporal rs-fMRI fluctuations of the two regions. For each subject, a 100×100 functional connectome matrix was thus generated, featuring the connections for every possible pair of brain regions.

2.3.4.2 Inter-site effects

The inter-site effects at a particular connection were defined as the absolute difference in average connectivity between two sites. In order to formally test the significance of the inter-site effects, we used a GLM including age, sex and residual motion as covariates (corrected to have a zero mean across subjects), as well as dummy variables coding for the average connectivity at each site. For each site, a "contrast" vector was coded to measure the difference in average connectivity between this site and the grand average

[11] `http://niak.simexp-lab.org/pipe_preprocessing.html`
[12] `https://github.com/SIMEXP/`

of functional connectivity combining all other sites. A p-value was generated for each connection to quantify the probability that the observed effect using this contrast was significantly different from zero (Worsley and Friston 1995). The number of false discovery was also controlled ($q = 0.05$) using a Benjamini-Hochberg false discovery rate (FDR) procedure (Benjamini and Hochberg 1995). To quantify the severity of inter-site effects, we derived Cohen's d effect size measure for each connection: $|\beta_c|/\hat{\sigma}$, with β_c being the weight associated with the contrast. The standard deviation from the noise $\hat{\sigma}$ was calculated as $\hat{\sigma} = \sqrt{\sum e^2/(N-K)}$, e being the residuals from the GLM, N the sample size and K the number of covariates in the model. As secondary analyses, t-tests were also implemented in the GLM to validate that age, sex as well as residual motion made significant contributions to the model.

2.3.5 Simulations

2.3.5.1 Data generation process

We implemented Monte-Carlo simulations to assess the detection sensitivity of group differences in rs-fMRI connectivity. The simulations were based on the 1000 FCP sample, with 8 sites totaling 345 subjects. The multisite simulations were sampled from 148 subjects, available across $S = 7$ sites. The monosite simulations were sampled from 195 subjects available at $S = 1$ site (Cambridge). For each simulation, a subset of subjects of a given size N was selected randomly and stratified by site. For each site, a ratio W of the selected subjects was randomly assigned to a so-called "patient" group. We focus our analysis on connections showing a fair-to-good test-retest reliability based on a previous study reporting 11 connections likely impacted by Alzheimer's disease, see Orban et al. (2015) for details. For each connection, a "pathology" effect was added to the connectivity measures of the subjects belonging to the "patient" group. This additive shift in connectivity for "patients" was selected as to achieve a specified effect size, defined below.

2.3.5.2 Effect size (Cohen's d)

The Cohen's d was used to quantify the effect size. For a group comparison, Cohen's d is defined as the difference μ between the means of the two groups, divided by the standard deviation of the measures within each group, here assumed to be equal. For a given connection between brain regions i and j, let $y_{i,j}$ be the functional connectivity measure for a particular subject of the 1000 FCP sample. If the subject was assigned to the "patient" group in a particular simulation, an effect was added to generate a simulated connectivity measure $y^*_{i,j}$ equal to $y_{i,j} + \mu$. For a specified effect size d, the parameter μ was set to $d \times s_{i,j}$, where $s_{i,j}$ is the standard deviation of connectivity between region i and j. The parameter $s_{i,j}$ was estimated as the standard deviation of connectivity measures across subjects in the mono-site sample (Cambridge), without any "pathological" effect simulated.

2.3.5.3 GLM tests

In order to detect changes between the simulated groups at each connection, a GLM was estimated from the simulated data, using age, sex and frame displacement as confounds (corrected to have a zero mean across subjects). To account for site-specific effects, $S-1$ dummy variables (binary vectors coding for each site) were added to the model, with S being the total number of sites used in the study, in addition to an intercept accounting for the global average. Finally, one dummy variable coded for the "patient" group. The regression coefficients of the linear model were estimated with ordinary least squares, and a t-test, with associated p-value, was calculated for the coefficient of the "patient" variable. A significant pathology effect was detected if the p value was smaller than a prescribed α level. The α level needs to be adjusted for multiple comparisons (in our case 11 connections, but this would depend on the number of connections selected in a particular study), which can be done in an adaptive manner using FDR. When connections are pre-specified, such as in e.g. Wang et al. (2012), a more liberal threshold can be applied. In our case, since we wanted to have a constant behavior independent of the effect size, we tested different typical values for α in $\{0.001, 0.01, 0.05\}$. For each

simulation sample b and each connection, we derived a p-value $p^{(*b)}$, and the effect was deemed detected if p^{*b} was less than α. The sensitivity of the test for a particular connection was evaluated by the frequency of positive detections over all simulation samples.

2.3.5.4 Prediction accuracy

In addition to mass univariate GLM tests, we also investigated a linear SVM (Cortes and Vapnik 1995) using a Monte-Carlo simulation of the prediction of clinical labels based on cross-validation. For SVM simulations, all possible connections between the 100 brain regions were used simultaneously to predict the presence of the simulated pathology in a given subject. For a participant assigned to the "patient" group, a "pathology" effect was only simulated in a set percentage of connections, which were randomly selected. The proportion of connections with a non-null effect was denoted as π_1. For a given simulation at sample size N, the SVM model was trained on N subjects selected randomly and stratified by site. The accuracy of the model was evaluated on a separate sample consisting of the remaining subjects, unused during training. For example, for a multisite simulation with $N = 80$ subjects for training, the model accuracy was tested on the remaining 68 subjects: 148 (available subjects) minus 80 (subjects in the training set). During training, a 10-fold cross-validation was used to optimize the hyper-parameters of the SVM independently for each simulation. The mean and standard deviation of accuracy scores across all samples were derived for each simulation scenario.

2.3.5.5 Simulation experiments

All the simulation parameters have been summarized below:

- Sample size N.
- Patient allocation ratio W.
- Number of sites S.

- The type of detection method, either GLM or SVM.
- For GLM tests, the false-positive rate α.
- For SVM tests, the proportion of "pathological" connections π_1.
- The effect size d.

For a given set of simulation parameters, we generated $B = 10^3$ Monte-Carlo samples to estimate either the sensitivity (for GLM test) or the accuracy (for SVM prediction) of the method. For all experiments, we investigated effect sizes $d \in \{0, 2\}$ with a step of 0.01 and $\alpha \in \{0.001, 0.01, 0.05\}$. The number of site(s) was $S = 1$ for the monosite analysis and $S = 7$ for the multisite analysis. We implemented the following experiments:

- $(\mathscr{E}_1)$ Test the impact of the sample size on GLM $N \in \{40, 80, 120\}$, with a fixed allocation ratio $W = 0.5$.
- $(\mathscr{E}_2)$ Test the impact of the allocation ratio on GLM $W \in \{0.5, 0.3, 0.15\}$ for a fixed sample size $N = 120$.
- $(\mathscr{E}_3)$ Test the impact of multisite correction (regressing out the site effects using dummy variables coding for each site) and affected connection volume (π_1) on the prediction accuracy. For the prediction scenario, we used a range of $\pi_1 \in \{0.1, 1, 5\%\}$, and two sample sizes $N \in \{80, 120\}$ subjects for training, with model accuracy estimated on $N = 68$ and $N = 28$, respectively.

2.4 Results

2.4.1 Inter-site effects in fMRI connectivity

2.4.1.1 Site effects in the default-mode network

We first focused on the connections associated with a seed region located in the posterior cingulate cortex, a key node of the default-mode network (DMN), which is one of the most widely studied resting-state networks (Greicius et al. 2004). The connections

Figure 2.1: Panel A: map of the DMN obtained using a seed in the posterior cingulate cortex, averaging all subjects and sites together (first row) and then averaging all subjects for each of the 8 sites (subsequent rows). Panel B shows the number of sites with a significant inter-site difference for each brain region (first row) and the significant differences between the average functional connectivity maps of one site versus all the others (subsequent rows).

were based on the Cambridge 100 parcellation, and were represented as a connectivity map, (Figure 2.1). Figure 2.1A shows the posterior cingulate cortex connectivity map, averaged across all subjects and all sites. The key regions of the DMN are easily identifiable, and include the posterior cingulate cortex, precuneus, inferior parietal lobule, anterior cingulate cortex, medial pre-frontal cortex (dorsal, anterior and ventral), superior frontal gyri and the medial temporal lobe (Damoiseaux et al. 2006, Dansereau et al. 2014, Yan et al. 2013b). The average connectivity map of the DMN was then extracted for each site, Figure 2.1A. Qualitatively, the DMN maps were consistent across sites, as expected based on the literature. We then tested for the significance of the site effects (Figure 2.1B), i.e. the difference in average connectivity at a given site and the average connectivity at all remaining sites. The statistical maps were corrected for multiple comparisons across the brain with FDR at $q \leq 0.05$ (Benjamini and Hochberg 1995). A significant site effect for at least one connection could be identified for every site, without exception, Figure 2.1B. Figure 2.1C shows how reproducible the significant site effects were in connectivity across the brain and sites. The identified significant connections were quite variable across sites, most of them being identified at less than three sites.

2.4.1.2 Site effects across the connectome

In order to extend these observations outside of the DMN, we derived the entire connectome using the Cambridge 100 parcellation. Figure 2.2A shows the average connectome, pooling all subjects and sites together. The regions have been re-ordered based on a hierarchical clustering with Ward criterion. A network structure is clearly visible as squares of high connectivity on the diagonal of the connectome (as outlined by black lines). Each diagonal square corresponds to the intra-network connectivity for a partition into 7 networks (Figure 2.2A). These 7 networks[13] were consistent with the major resting-state networks reported using a cluster analysis in previous works (e.g. Bellec et al. 2010a, Power et al. 2011, van den Heuvel et al. 2008, Yeo et al. 2011): the DMN, visual, sensorimotor, dorsal and ventral attentional networks, mesolimbic and cerebellar

[13] `http://neurovault.org/images/39184/`

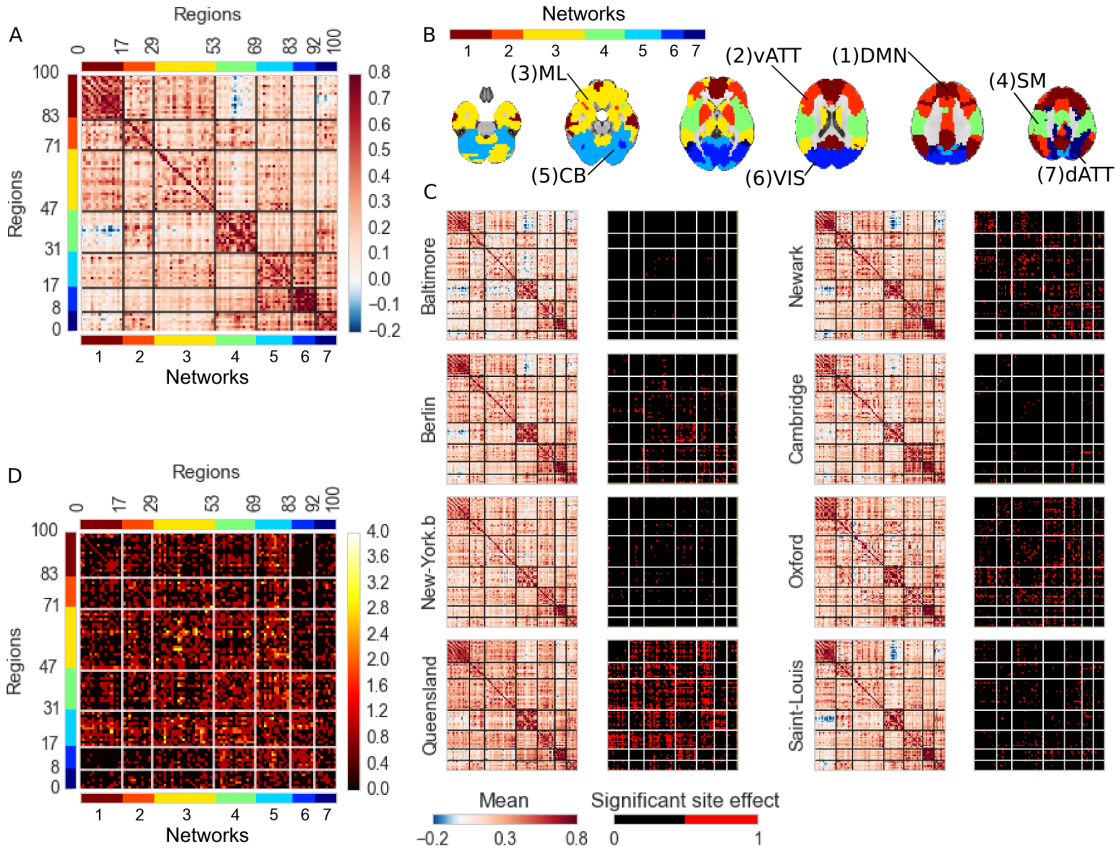

Figure 2.2: Panel A shows the average functional connectomes for 8 sites of the 1000 FCP. Colors next to the *x* and *y* axis correspond to different networks in a 7-cluster solution of the matrix, obtained from a hierarchical clustering (Ward criterion). Panel B presents the corresponding 7 brain networks, along with labels. Panel C shows average connectomes for individual sites, as well as connections with a significant site effect. Panel D shows the number of sites at which a given connection was detected as significant. ML: mesolimbic, CB: cerebellar, VIS: visual, vATT: ventral attentional, dATT: dorsal attentional, DMN: default mode network, SM: sensorimotor.

networks were identified (Figure 2.2B). Figure 2.2C shows how this large-scale connectome organization varied from site to site. The average connectivity per site as well as significant differences with the average of the remaining sites ($q \leq 0.05$) is shown in Figure 2.2C. Visually, consistent with our previous observations in the DMN, the organization of the average connectome into large-scale resting-state networks was preserved across all sites.

Some significant site effects were still detected in the connectivity both within each network, as well as between networks. By counting the number of sites showing a significant effect for each pair of regions, it was apparent that significant site effects were quite variable in their localization and spread across the full connectome (Figure 2.2D). Concerning the association with the other confounding variables in the model (sex, age and motion) many connections were found to be significantly associated with motion, see Supplementary Material Figure S5, although very few connections were found to be significantly associated with the sex and age, see Supplementary Material Figure S6 and S7. We also checked that the analysis was not predominantly driven by the larger Cambridge site. We thus ran the same analysis excluding that site (see Supplementary Material Figure S8). The number of significant pairs remained very similar, although the spatial location of half of the significant connectivity pairs changed when the large Cambridge site was removed from the analysis. Those findings do not qualitatively change our conclusion, but they influence the location of the significant connections. These differences may be due to the intrinsic variability in the statistical test, and not just the size of the Cambridge site. In summary, those findings support the inclusion of age, sex and motion parameters in a GLM in order to remove their confounding effects in addition to site effects.

2.4.1.3 Site effects vs. within-site variations across subjects

We measured the amplitude of inter-site effects, represented as violin plots across connections using either the absolute difference in average connectivity (Figure 2.3A,C) or Cohen's d effect size measures (Figure 2.3B,D). The violin plots include either every connection from the BASC Cambridge parcellation (Figure 2.3A,B), or only the 11

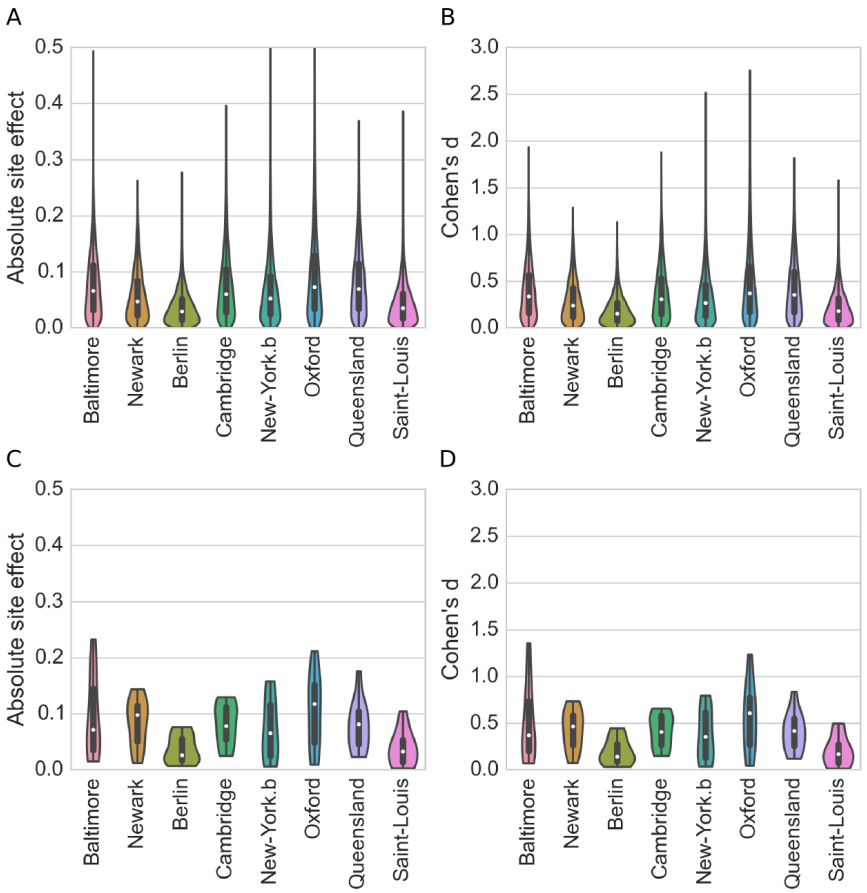

Figure 2.3: Effect size of the inter-site effects from a subset of 8 sites from the 1000 FCP. Panels A,C show the distribution of absolute differences in functional connectivity, while panels B,D show Cohen's *d* measures of inter-site effects. Panels A,B show violin plots across every connections in the BASC Cambridge 100 parcellation, while Panels C,D focus on the selected 11 functional connections used in simulations, only.

connections selected for Monte-Carlo simulations (Figure 2.3C,D). For absolute differences, the distributions were mostly consistent across sites, with a median around 0.06, 5% percentile near 0 and 95% percentiles in the 0.08- 0.1 range. For Cohen's d, the distributions were also consistent across sites, with a median around 0.33, 5% percentile near 0 and 95% percentiles in the 0.4- 0.6 range. These effect sizes are typically deemed small-to-moderate (Cohen 1992), although such a qualitative assessment needs to be refined based on each application. This result thus suggests that the impact of additive inter-site effects on statistical tests will be limited. Similar findings were observed across all possible connections, or across the 11 pairs of connections selected in the simulation study.

2.4.1.4 Differences in standard deviation across sites

We also investigated the site differences in standard deviation of connectivity across subjects, see Supplementary Figure S1 for the DMN, Supplementary Material S2 for the connectomes. The standard group GLM assumes equal variance of resting-state connectivity across all subjects, or "homoscedasticity". Significant differences in across-subject standard deviation between sites violates the homoscedastic assumption, and may jeopardize the validity of the false-positive rates of the model. Qualitatively, we first observed that the sites showing the larger number of differences were the one with the most temporal variance among connections see Supplementary Figure S3. We then ran a White's test aimed at rejecting homoscedasticity at each connection, independently. The White's tests resulted in a family of p-values, which was corrected for multiple comparisons using FDR ($q < 0.05$). The homoscedastic hypothesis was rejected in a large portion of connections. This was expected due to the large overall number of subjects and consequently large statistical power of White's procedure. However, despite reaching significance, the absolute difference in the average standard deviation between two sites was 19% of the grand average standard deviation, on average across pairs of sites. Such a small departure from homoscedasticity likely has only a mild impact on the GLM, which we formally investigated using Monte-Carlo simulations.

2.4.2 Multisite Monte-Carlo simulations

2.4.2.1 Validity of the control of false positives in the GLM

An excellent control of the false positive rate was observed at all nominal levels $\alpha \in \{0.001, 0.01, 0.05\}$, both in monosite simulations or in multisite simulations, when site covariates were included in the GLM, see Figure 2.4. This means that the nominal, user-specified, false positive rate matched precisely with the effective false positive rate measured in the simulations. This observation held for any combination of allocation ratio, $W \in \{15\%, 30\%, 50\%\}$, and sample size, $N \in \{40, 80, 120\}$. By contrast, when no site covariates were included in the GLM, the false positive rate was not controlled appropriately, sometimes by a wide margin. In the absence of site covariates, the procedure was sometimes too conservative, e.g. $W = 50\%$, and sometimes very liberal, e.g. $N = 120, W = 15\%$. This experiment showed that, despite the mild departure from homoscedasticity reported above, the GLM does control for false-positive rate at each connection very precisely, if and only if site covariates are included in the model.

2.4.2.2 Statistical power and effect size

Figure 2.5A shows the relationship between effect size and a GLM detection power in experiment ($\mathscr{E}_1$), i.e. for a fixed allocation ratio ($W = 50\%$) and three different sample sizes, $N \in \{40, 80, 120\}$. The average and std of detection power was plotted across the 11 selected connections. The variations of statistical power across connections were very small for monosite simulations, as the effect size was adjusted based on the standard deviation of each connection within that sample. As expected, the sensitivity increased with sample size, quite markedly. In multisite simulations ($S = 7$), for a large effect size ($d = 1$), the detection power was 20% with 40 subjects , 80% with 80 subjects and 95% with 120 subjects. The sensitivity was larger with a single site than a multisite sample, yet the difference between the two decreased as sample size increased. With $N = 40$ and $d = 1$, the detection power was close to 30% for a single site sample, compared to 20% for the multisite sample. With $N = 120$ and $d = 1$, the difference in sensitivity was only of a few percent. The same trend was apparent for all tested effect sizes as well as for

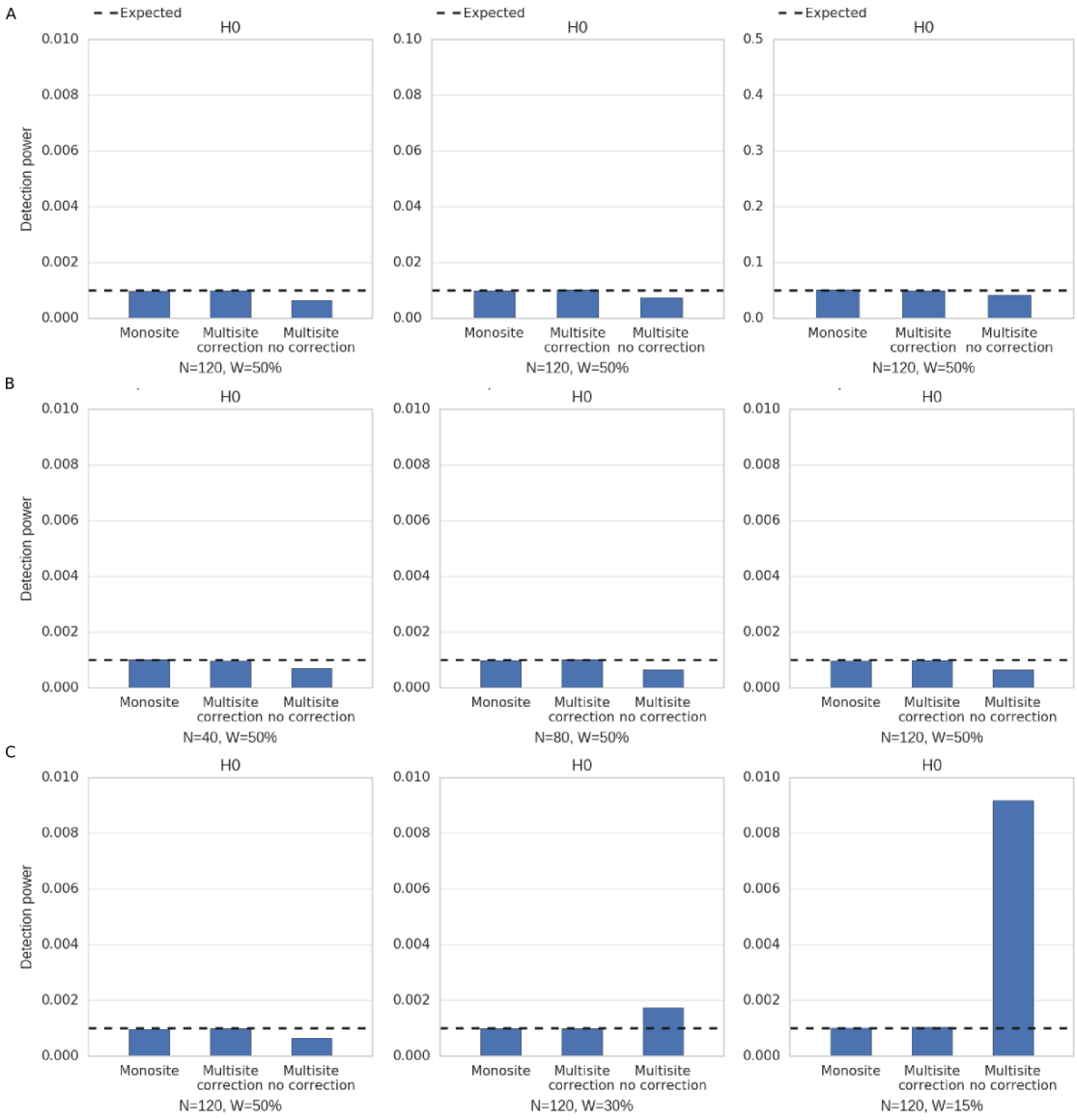

Figure 2.4: Monte-Carlo simulation of the false positive rate in the absence of group differences ($d = 0$), either for a monosite ($S = 1$, left), a multisite ($S = 7$) with (middle) or without (right) site covariates included in the GLM. In panel A, three different α values have been tested, $\alpha \in \{0.001, 0.01, 0.05\}$ with a fixed sample size and patient allocation ratio ($N = 120, W = 50\%$). In panel B, three different sample sizes have been tested, $N \in \{40, 80, 120\}$ with a fixed patient allocation ratio ($W = 50\%$) (Experiment ($\mathscr{E}_1$)). In panel C, three different patient allocation ratios have been tested, $W \in \{50\%, 30\%, 15\%\}$ with a fixed sample size ($N = 120$) (Experiment ($\mathscr{E}_2$)).

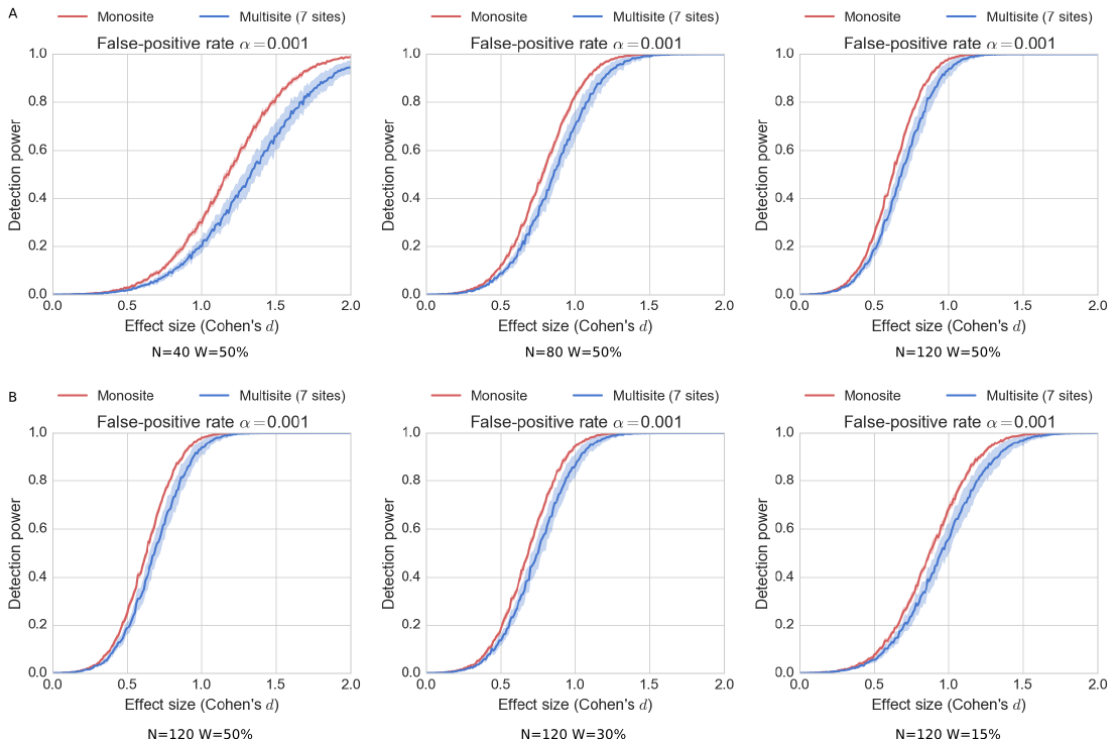

Figure 2.5: Monte-Carlo simulation of detection power as a function of the effect size $d \in [0,2]$, either for a monosite ($S = 1$, in red) or a multisite ($S = 7$, in blue) sample, when testing differences between two groups with a GLM and a false-positive rate $\alpha = 0.001$. The plain curves are the average statistical power across 11 connections, and the shaded area represents ± 1 standard deviation across connections. In panel A, the patient allocation ratio is fixed ($W = 50\%$) and three different sample sizes have been tested, $N \in \{40, 80, 120\}$ (Experiment $(\mathscr{E}_1)$). In panel B, the sample size is fixed ($N = 120$) and three different patient allocation ratios have been tested $W \in \{15\%, 30\%, 50\%\}$ (Experiment $(\mathscr{E}_2)$).

$\alpha \in \{0.01, 0.05\}$ (not shown).

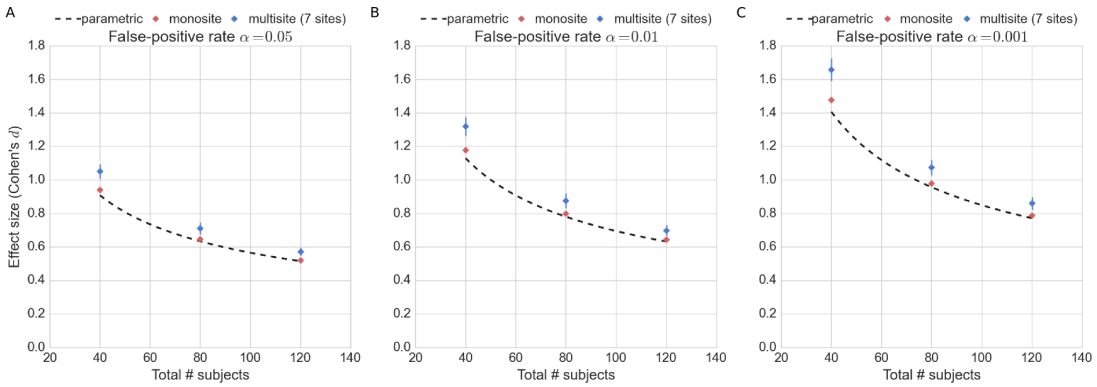

Figure 2.6: Effect size detectable at 80% sensitivity as a function of sample size, for different false-positive rate $\alpha \in \{0.05, 0.01, 0.001\}$ (experiment $(\mathscr{E}_1)$). All simulations used a balanced patient allocation ratio $W = 50\%$. The monosite performance is shown in red and the multisite in blue. The dotted black line shows the detectable effect size for a classical parametric t-test.

2.4.2.3 Statistical power and group allocation ratio

Figure 2.5B shows the relationship between effect size and a GLM detection power in experiment $(\mathscr{E}_2)$, i.e. for a fixed sample size ($N = 120$) and three different patient allocation ratio, $W \in \{15\%, 30\%, 50\%\}$. Overall, we found that the detection power increased with W. For example, with $d = 1$, the detection power was 65% for $W = 15\%$, and increased to 90% with $W = 30\%$, and finally 95% for $W = 50\%$. The impact of W was observed in both monosite and multisite samples, with an optimal allocation ratio of $W = 50\%$ for both. This observation was also made for $\alpha \in \{0.01, 0.05\}$ (not shown).

2.4.2.4 Detectable effect size, as a function of sample size

An alternative summary of experiment $(\mathscr{E}_1)$ is to represent the effect size that can be detected with 80% sensitivity, as a function of sample size for monosite and multisite configurations, see Figure 2.6. As a reference, we computed the same curve for parametric t-test comparisons, under assumptions of normality. As expected, the detectable effect size for parametric t-tests closely followed the monosite estimation. For a small

sample size ($N = 40$), the detectable effect size was notably larger in multisite configurations than in a monosite configuration (difference of about 0.25 in Cohen's d for $\alpha = 0.001$). However, the difference decreased for large sample sizes to become smaller than 0.1 with $N = 120$ and $\alpha = 0.001$. The lowest detectable effect size for a sensitivity of 80% at $\alpha = 0.05$ was about $d = 0.8$, achieved in a monosite configuration with $N = 120$. At this sample size, the difference between single and multisite configurations was marginal, with only a few percent's of difference in detectable effect sizes.

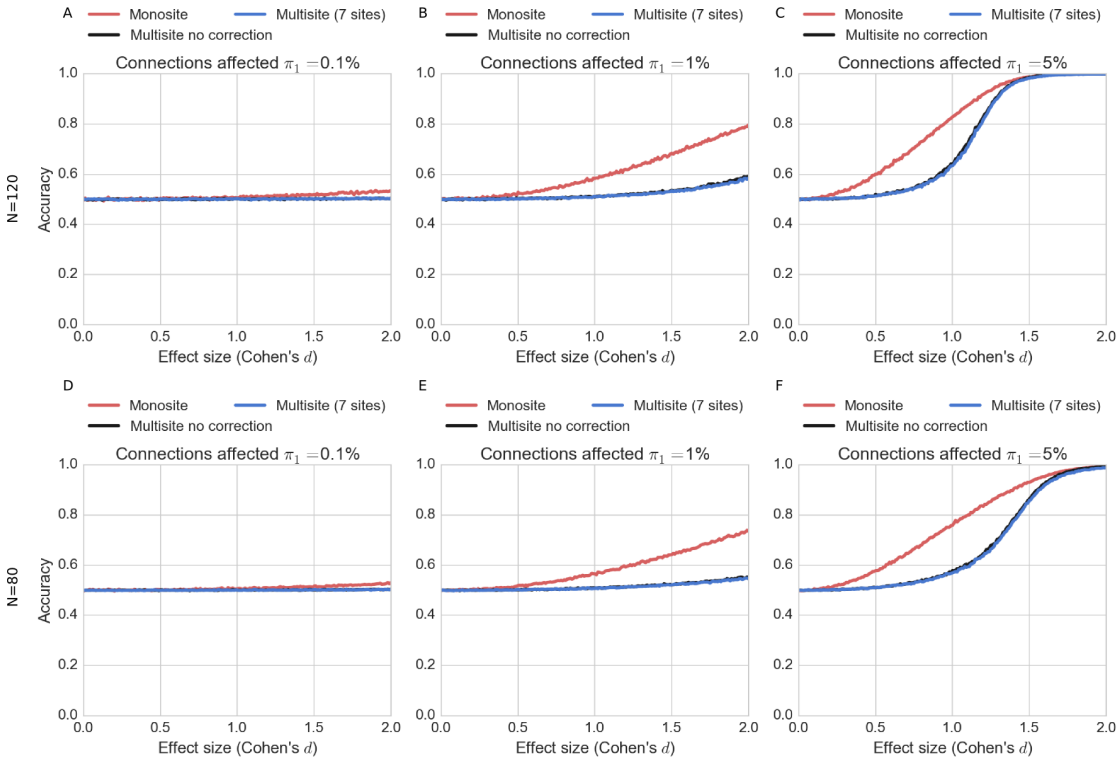

Figure 2.7: Prediction accuracy of patient vs. controls as a function of effect size. Three simulation settings are presented on each plot: monosite (red curve), multisite with regression of site effects ($S = 7$, blue curve), and multisite without regression of site effects ($S = 7$, black curve). Accuracy was estimated over $B = 10^3$ simulation samples with a patient allocation ratio $W = 50\%$ and 3 volumes of affected connections $\pi_1 = 0.1\%$ (left column), $\pi_1 = 1\%$ (middle column) and $\pi_1 = 5\%$ (right column). Two sample sizes were tested: $N = 120$ randomly selected subjects for training, with the remaining $N = 28$ to estimate accuracy (first row), and $N = 80$ randomly selected subjects for training, with the remaining $N = 68$ to estimate accuracy (second row).

2.4.2.5 Prediction accuracy

In experiment ($\mathscr{E}_3$), we examined the impact of effect size and the volume of affected connections on prediction accuracy in a SVM, see Figure 2.7. The volume of changes π_1 had a major impact on prediction accuracy. At $\pi_1 = 0.1\%$ (around 5 connections) the accuracy level was at chance level across all tested effect sizes, (Figure 2.7A). With $\pi_1 = 1\%$, accuracy slightly increased, but effect sizes larger than $d = 2$ were still required to reach over 80% accuracy (Figure 2.7B). With $\pi_1 = 5\%$, 95% accuracy was achieved at the same effect size (about $d = 1.5$) for monosite and multisite simulations, although the accuracy in multisite simulations was notably lower than for monosite simulations across most effect sizes (Figure 2.7C). The relationship between effect size and accuracy followed a sigmoidal curve in both settings, yet a sharper, and later transition between very low and very high accuracy was observed in multisite simulations. Interestingly, correcting for site effects by regressing out the dummy variable before running the SVM classifier had no impact on accuracy levels. The sample size ($N = 80$ vs $N = 120$ for training) did have a moderate effect on prediction accuracy: for $\pi_1 = 5\%$ and $d = 1$ and monosite simulations, accuracy was about 85% with $N = 120$ (Figure 2.7C) and 75% with $N = 80$ (Figure 2.7F).

2.5 Discussion and conclusions

2.5.1 Inter-site effects in rs-fMRI connectivity

Typical resting-state networks, such as the DMN, the attentional, visual and sensorimotor networks, were reliably found across sites. This was strongly expected given the relative consistency of their distribution across individuals, studies, preprocessing approaches or even methods used to extract networks (e.g. Bellec et al. 2010b, Damoiseaux et al. 2006, Power et al. 2011, van den Heuvel et al. 2008, Yeo et al. 2011). We however found that significant differences in average connectivity existed between sites, as previously reported by Yan et al. (2013b). These site effects in connectivity may undermine the generalization of the results derived at a single site. The inter-subject

(intra-site) standard deviation of the connections was found to be more than twice as large as the inter-site absolute effect, on average across brain connections. This effect size measured in Cohen's d would be deemed small-to-moderate, which suggests that the impact of additive inter-site effects on statistical tests will be limited. This is a reassuring finding supporting the feasibility of statistical tests pooling fMRI data across multiple sites. Previous studies (Brown et al. 2011, Sutton et al. 2008) had reported inter-site variance up to 10 times smaller than inter-subject variability, but these studies had much more homogeneous scanning environments than ours and also used different fMRI outcome measures. In our case, we still investigated only 3T scanners, mostly Siemens, and inter-site effects may be larger when considering other manufacturers or field strengths.

2.5.2 Statistical power and multisite rs-fMRI

After accounting for site-related additive effects in a GLM, the multisite simulation pooling 7 sites together showed detection power close to that of a monosite simulation with equivalent sample size. The difference was noticeable for small sample size (total $N = 40$), and became very small for a sample size $N = 120$. Another observation was that, for a given detection power, the lowest effect size that we were able to detect was more variable across connections for a low sample size. We demonstrated that a parametric group GLM does control precisely for the rate of false positive discoveries, even in multisite settings, as long as site covariates are included in the model. Taken together, these observations suggest to use sample sizes larger than 100 subjects for GLM multisite studies. This conclusion may depend on the number of sites pooled in the study and the actual number of subjects in each of those sites, which we could not test in this work due to the size of the available sample.

2.5.3 Modeling site effects as random variables

We modeled the effect of each site on the average connectivity between any given pair of regions as a fixed effect. This means that the proposed GLM inference does apply only to collection of sites included in a given analysis. The linear mixed-effects model (Chen et al. 2013) would allow more powerful inferences: by modeling site effects as random variables, following a specific distribution (e.g. Gaussian), we would be able to generalize observations potentially to any collection of sites, provided our assumptions are accurate. The sample of sites available for this study (7 at most) is however too small in our view to correctly estimate the variability of effects across sites. This work would also require to formulate and investigate empirically as well as on simulations different models for the distribution of inter-site variations of site effects (e.g. Gaussian distribution).

2.5.4 Site heteroscedasticity

We observed mild heteroscedasticity across sites. Our simulations showed that this does not compromise the control of false positive rate in the GLM, even under homoscedastic assumptions, with the range of contrasts we investigated. Regression models more robust to heteroscedasticity may be investigated in the future, e.g. weighted least squares regression or linear mixed-effects modeling (Chen et al. 2013).

2.5.5 Statistical power and sample size

For a medium effect size, e.g. $d = 0.5$, the sensitivity was low (below 20%), even for monosite simulations with $N = 120$ subjects. This sobering result supports the current trend in the literature to pool multiple data samples to increase sample size, at the cost of decreased homogeneity. We also found that resting-state studies based on 40 subjects or less, even at a single site, are seriously underpowered, except for extremely large effect sizes (Cohen's d greater than 1.5). Finally, unbalanced patient allocation ratio in site samples greatly reduces sensitivity, even in monosite studies. Balanced datasets, i.e. with equal numbers of patients and controls at each site, should therefore be favored.

2.5.6 Prediction

Comparing the monosite and the multisite accuracy curves reveals a substantial drop in accuracy from monosite to multisite across a broad range of effect sizes. However, it should be noted that classifiers trained across multiple data sources will likely generalize better to new observations, which is likely a critical feature in most applications and reflects the true potential clinical utility of this type of technique. Our conclusions are consistent with the work of Nielsen et al. (2013), which compares the prediction of a clinical diagnosis of autism in monosite vs. multisite settings. The authors concluded that the prediction accuracy for the multisite sample was significantly smaller than for the monosite sample. A somewhat surprising observation in our analysis was that linear correction for site-specific effects did not improve accuracy of prediction using SVM. The SVM model seems to learn features that are invariant across sites, maybe focusing on connections with the smallest site effect, or looking at differences between connections similarly impacted by a site effect. Finally, an important conclusion of our simulations was that the volume of brain connections affected by a disease impacts accuracy as much as the effect size per connection. This suggests that feature reduction and/or selection is a very important step to improve sensitivity to small effect sizes.

2.5.7 Beyond additive site effect

An important limitation to our study is that we only investigated the impact of additive effects in brain connectivity across sites. Areas of future work include interactions between site effects and pathology, possibly in the form of polynomial and non-linear interactions. We hope that, in the future, fMRI data acquired on clinical cohorts at tens of sites will become available, which will enable researchers to test empirically the presence of such interaction effects.

2.5.8 Other types of multisite data

Another limitation of our study is that we only investigated multisite data featuring roughly equal sample sizes with fairly balanced patient allocation ratios at each site.

Multisite studies including a very large number of sites with sometimes only a few subjects per site are however quite common, e.g. the Alzheimer's disease neuroimaging initiative (ADNI) (Mueller et al. 2005) and many pharmaceutical clinical trials at phase II and III [14]. In this type of design, the multisite effect may play a much more pronounced role than in our simulations as it cannot be modeled in the GLM, and will become an intrinsic added source of inter-subject variance (Feaster et al. 2011). Unfortunately, this type of design could not be tested with the current dataset due to the limited number of sites available. This represents an important avenue of future work.

2.5.9 Underlying causes of the site effects

Not all sites seemed to be equally impacted by the site effects, with sites like Berlin or Saint-Louis showing a small number of connections significantly different then the grand average connectivity matrix, while sites like Baltimore, Queensland and Oxford showed many more connections affected by the site effects. Interestingly this can potentially be due to temporal variance of the connections (see Supplementary Figure S3) partly explained by the scanner make since Queensland and Baltimore site used scanners from different makers (namely Bruker and Philips) than the rest of the sites used in this study (Siemens scanners). This may suggest that scanners SNR (signal to noise ratio) may partly explain the variance of connectivity. These differences may not be statistically significant, or they may reflect real differences due to protocol, scanner characteristics at these sites or differences in sampling across sites. Multiple causes may be interacting together to produce the site effects, as reported by Yan et al. (2013b), although some of these sources of variance could be better controlled like the scanner parameters, paired with the use of a phantom to promote more homogeneous configurations across sites (Friedman and Glover 2006, Friedman et al. 2006; 2008, Glover et al. 2012). Even in standardized experiments, it should be noted that differences in scanner protocols remain (Brown et al. 2011). A much larger multisite sample with systematically varying parameters could enable a data-driven identification of the critical parameters impacting

[14] `http://www.rochetrials.com/trialDetailsGet.action?studyNumber=BP28248`

site effects. The various releases made by the INDI initiative may fill that gap in the literature in the future, as the scanner protocols are much better described in recent releases, such as CoRR (Zuo et al. 2014), than they were in the initial FCP release. These findings stress the need for more work to find the source of that variance rather than ad-hoc procedures to correct for them.

2.6 Acknowledgments

Parts of this work were presented at the 2013 annual meeting of the Organization for Human Brain Mapping, as well as the 2013 Alzheimer's Association International Conference (AAIC) (Dansereau et al. 2013). The authors are grateful to the members of the 1000 functional connectome consortium for publicly releasing their datasets. The computational resources used to perform the data analysis were provided by Compute-Canada[15] and CLUMEQ[16], which is funded in part by NSERC (MRS), FQRNT, and McGill University. This project was funded by NSERC grant number RN000028 and the Canadian Consortium on Neurodegeneration in Aging (CCNA), through a grant from the Canadian Institute of Health Research and funding from several partners including SANOFI-ADVENTIS R&D. PB is supported by a salary award from "Fonds de recherche du Québec – Santé" and the Courtois Foundation.

[15] `https://computecanada.org/`
[16] `http://www.clumeq.mcgill.ca/`

Supplementary Material – Statistical power and prediction accuracy in multisite resting-state fMRI connectivity

Submitted to Neuroimage.

C. Dansereau[1,2], Y. Benhajali[1,3], C. Risterucci[4], E. Merlo Pich[4], P. Orban[1], D. Arnold[5], P. Bellec[1,2]

[1]Centre de Recherche de l'Institut Universitaire de Gériatrie de Montréal, Montréal, CA

[2]Department of Computer Science and Operations Research, University of Montreal, Montreal, CA

[3]Département d'anthropologie, Universitde Montréal, Montréal, CA

[4]Clinical Imaging, pRED, F.Hoffman-La Roche, Basel, CH

[5]NeuroRx inc., Montréal, CA

For all questions regarding the paper, please address correspondence to Pierre Bellec, CRIUGM, 4545 Queen Mary, Montreal, QC, H3W 1W5, Canada. Email: pierre.bellec (at) criugm.qc.ca.

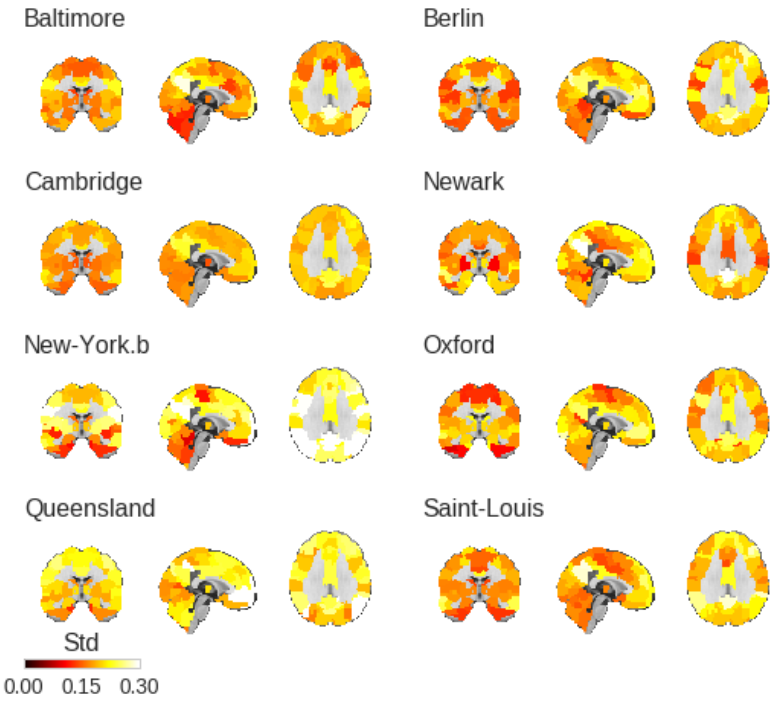

Figure S1: Standard deviation of resting-state connectivity across subjects, in the DMN, for each site, superimposed on the MNI152 template.

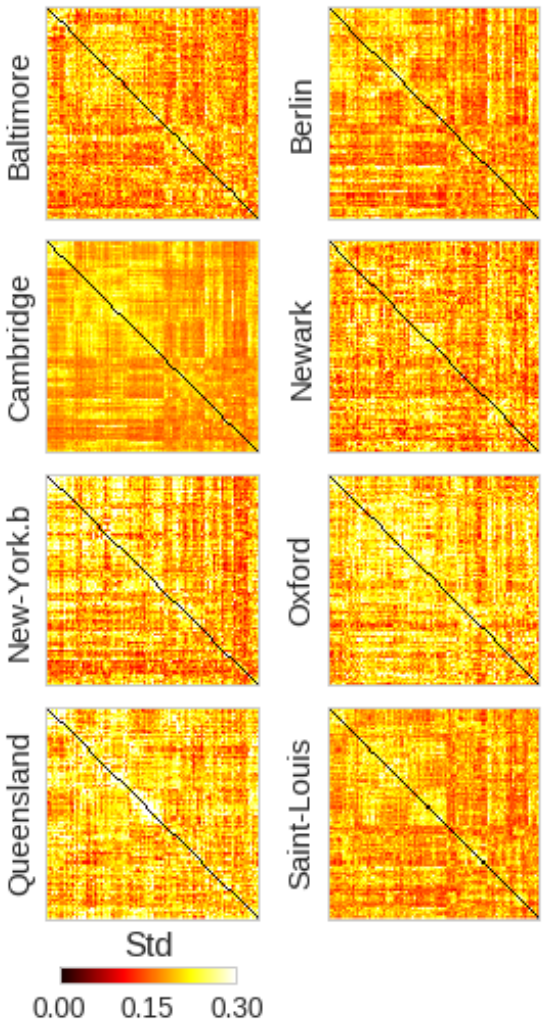

Figure S2: Standard deviation of resting-state connectivity across subjects, for the full connectome and each site.

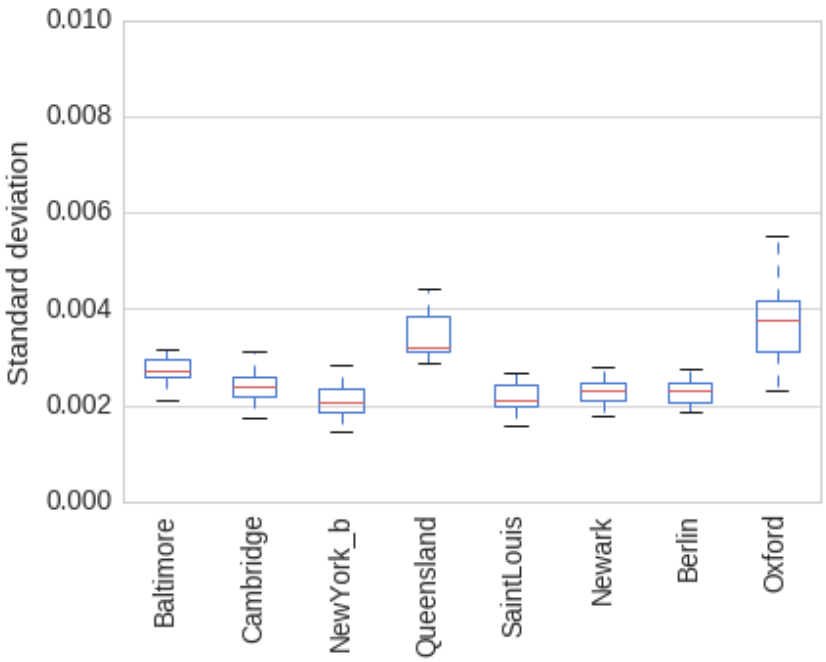

Figure S3: Standard deviation of resting-state time-series across subjects, averaged across all connections, at each site.

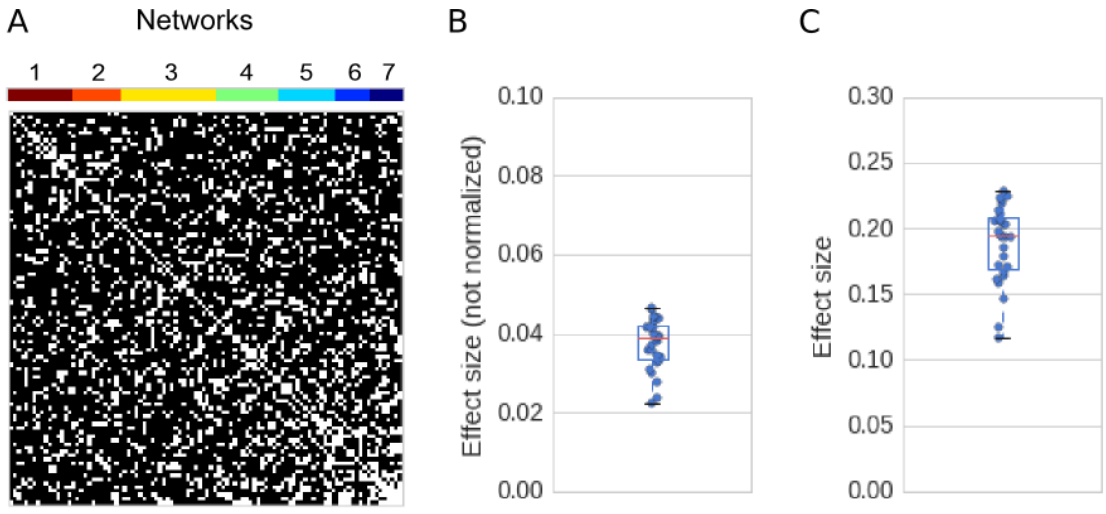

Figure S4: Panel A shows the results of a White test for homoscedasticity, across sites. Panel B show the average absolute difference in standard deviation between any pair of sites, and Panel C show the same difference, relative to the average of the standard deviation at the two sites.

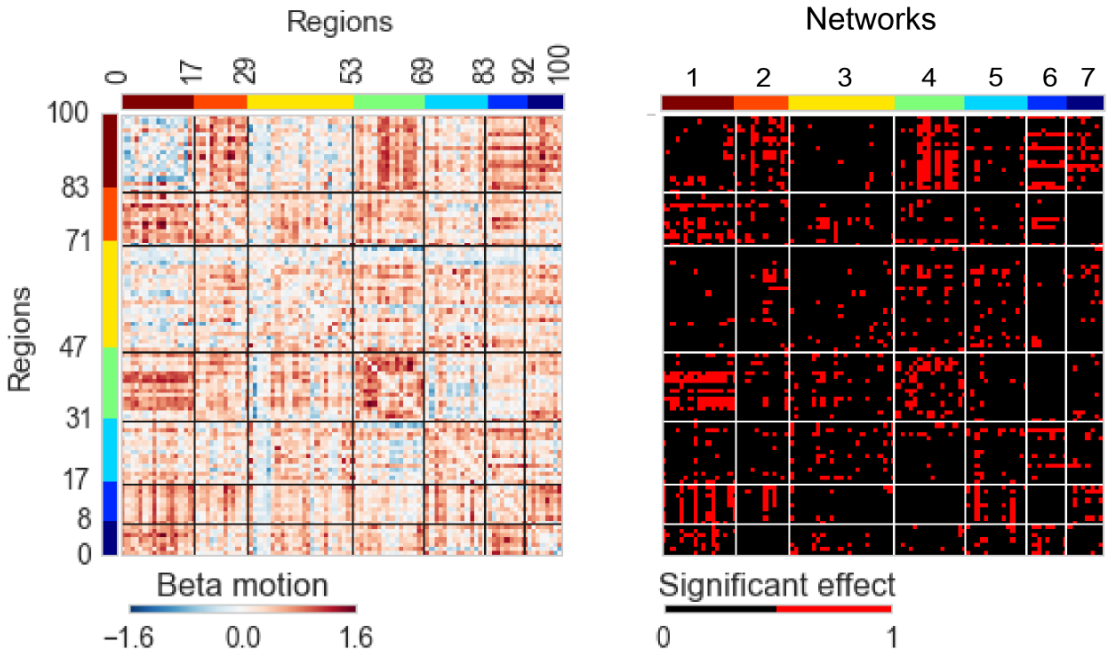

Figure S5: The figure shows average connectomes across all sites, as well as connections with a significant motion effect.

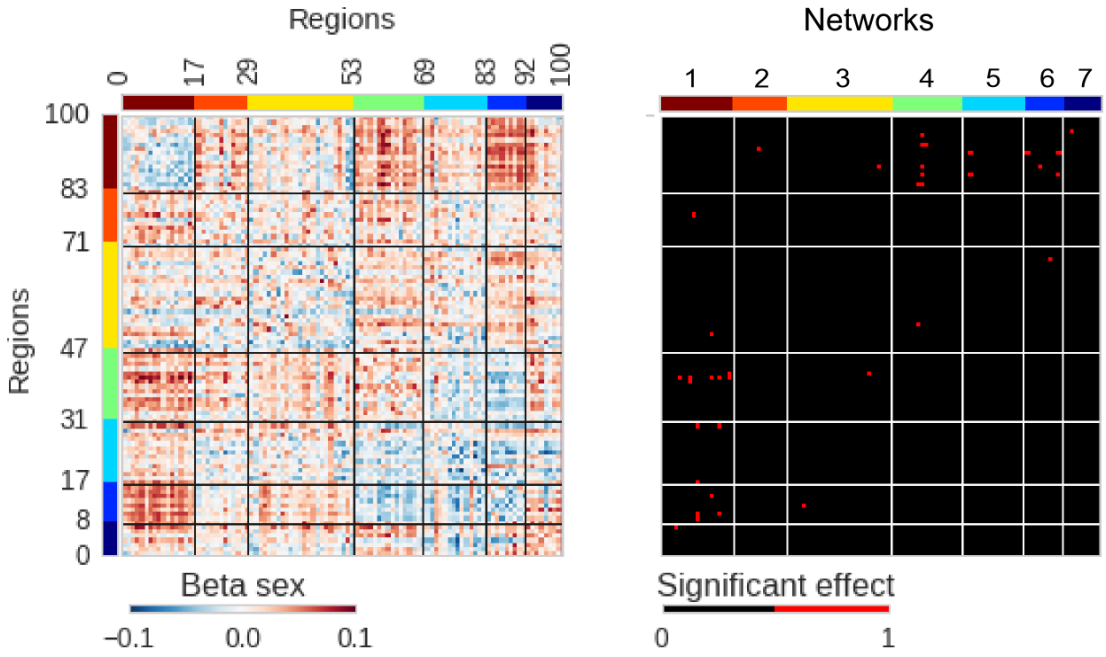

Figure S6: The figure shows average connectomes across all sites, as well as connections with a significant sex effect.

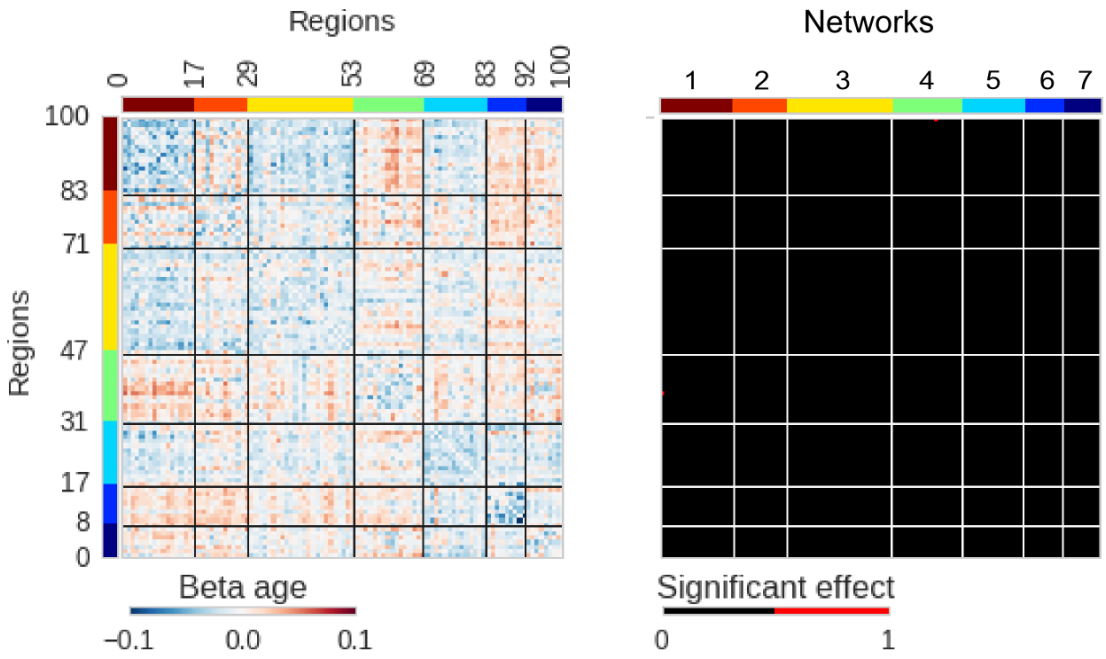

Figure S7: The figure shows average connectomes across all sites, as well as connections with a significant age effect.

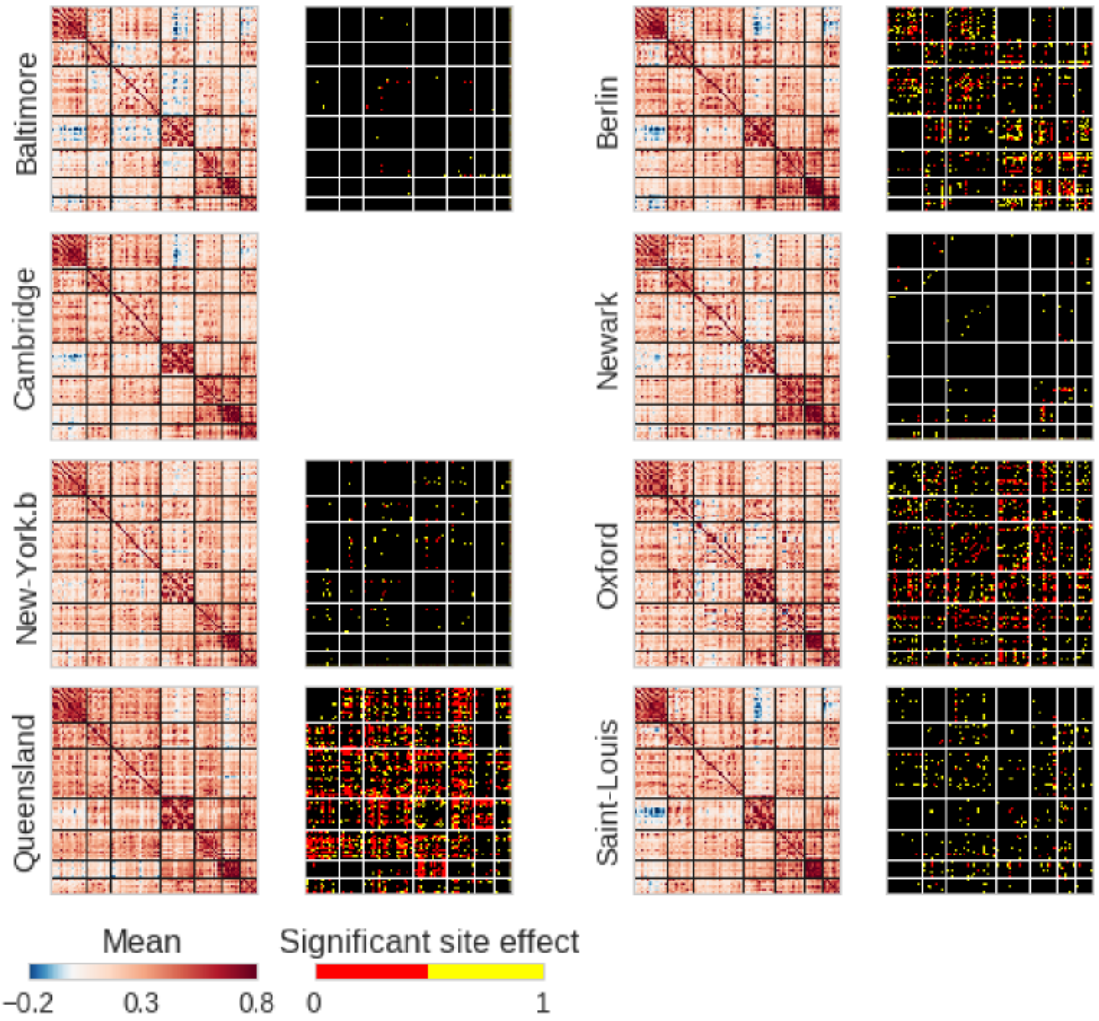

Figure S8: Average connectomes for individual sites, as well as connections with a significant site effect. This Figure is identical to Figure 2 in the paper, with the difference that the Cambridge site was excluded from the analysis. The intersection ($\cap$) of the significant site effects are shown in red and the symmetric difference ($\triangle$) of the significant site effects are shown in yellow. Baltimore $\cap: 9, \triangle: 16$, Berlin $\cap: 318, \triangle: 333$, Newark $\cap: 23, \triangle: 36$, New-York.b $\cap: 25, \triangle: 45$, Oxford $\cap: 377, \triangle: 251$, Queensland $\cap: 946, \triangle: 389$, Saint-Louis $\cap: 49, \triangle: 162$

CHAPTER 3

MULTISITE GENERALIZABILITY OF SCHIZOPHRENIA DIAGNOSIS CLASSIFICATION BASED ON FUNCTIONAL BRAIN CONNECTIVITY

Published in Schizophrenia Research. 2017[1]

P. Orban†, C. Dansereau†, L. Desbois, V. Mongeau-Pérusse, C. Giguère, H. Nguyen, A. Mendrek, E. Stip, P. Bellec

† Equally contributed to this work.

3.1 Abstract

Our objective was to assess the generalizability, across sites and cognitive contexts, of schizophrenia classification based on functional brain connectivity. We tested different training-test scenarios combining fMRI data from 191 schizophrenia patients and 191 matched healthy controls obtained at 6 scanning sites and under different task conditions. Diagnosis classification accuracy generalized well to a novel site and cognitive context provided data from multiple sites were used for classifier training. By contrast, lower classification accuracy was achieved when data from a single distinct site was used for training. These findings indicate that it is beneficial to use multisite data to train fMRI-based classifiers intended for large-scale use in the clinical realm.

3.2 Introduction

Psychiatrists and other mental health professionals could benefit in the not-so-far future from neuroimaging-based classification tools to assist diagnosis and prognosis in mental illness (Huys et al. 2016). Recent developments in the neuroimaging field

[1] http://dx.doi.org/10.1016/j.schres.2017.05.027

have led to a shift from group comparisons based on averaging across subjects to machine learning techniques making prediction at the individual level (Dubois and Adolphs 2016). In this approach, the emphasis is put on the ability of an algorithm to classify individuals into clinical categories with good generalizability to unseen subjects. Over the last decade, hundreds of studies have successfully classified various psychiatric and neurological disorders based on in vivo brain imaging (reviewed in Arbabshirani et al. (2017), Wolfers et al. (2015)). For instance, Arbabshirani et al. (2017) identified 30 published studies that distinguished schizophrenia patients from healthy controls with an average accuracy of 83% using functional magnetic resonance imaging (fMRI), either under task or rest states.

To date, however, the vast majority of classification works in mental illness were performed in a research context, using data from single sites of acquisition. Such findings may not generalize to large-scale clinical settings, with patients being scanned at widely-spread sites and possibly under various mental states. In most cases, the performance of classifiers was only assessed for unseen, test subjects with the exact same characteristics as the sample used for training. Yet, using gender as a proof-of-concept target variable, there was initial evidence that classifiers only poorly generalize to data drawn from other site samples (Huf et al., 2014). The inclusion of data from multiple sites during training improved the classifier performance for data of unseen sites.

In schizophrenia, a study pooling fMRI data from two distinct scanning sites reported similar prediction accuracy levels irrespective of whether test data were drawn from the dataset used for training or not, thus suggesting good generalizability (Skåtun et al. 2016). However, this result appears at odds with a recent fMRI study in autism that showed poorer accuracy for inter-site than intra-site training/test configurations, depending on the ratio of training set used (Abraham et al. 2016). In the case of inter-site testing, data pooled from 4 sites were used for training the classifier, which was tested on data from a fifth site. Yet, none of these two studies specifically evaluated whether using multisite training data could compensate to some extent for the deleterious effect of inter-site testing, by assuming the actual presence of such an effect. In the present work, we sought to address this question based on fMRI brain connectivity in schizophrenia.

Since it is impossible to completely control the variations in mental states in realistic clinical situations, we further promoted the complexity of the classification problem by including data obtained in distinct cognitive task conditions across sites. Mass univariate findings have indicated that cognitive state does not further impact on the nature of functional brain connectivity alterations in schizophrenia (Kaufmann et al. 2017, Orban et al. 2017b). However, the potential influence of cognitive context on classification performance in a multivariate analysis should not be rejected.

3.3 Method

3.3.1 Datasets

Brain imaging data from 6 independent studies were obtained through either the SchizConnect and OpenfMRI data sharing platforms[2][3] or local scanning (Çetin et al. 2014, Gollub et al. 2013, Kogan et al. 2016, Orban et al. 2017b, Poldrack et al. 2016, Wang et al. 2016). The 6 datasets differed in terms of both scanning site and cognitive context during fMRI data acquisition (resting-state, emotional memory, Sternberg item recognition paradigm, N-back, task-switching and oddball tasks). Classification analyses included fMRI data from 382 subjects, 191 patients diagnosed with schizophrenia and 191 healthy controls. Subjects provided informed consent to participate in their respective studies and ethics approval was obtained at the site of secondary analysis (Centre de Recherche de l'Institut Universitaire de Gériatrie de Montréal, Montréal, Canada).

3.3.2 Subjects matching

Sample size differed between sites (N = 84, 82, 70, 62, 50 and 34). Site samples were obtained after subjects were selected in order to ensure even proportions of schizophrenia patients and controls within each site (N = 42, 41, 35, 31, 25 and 17 subjects per group) and to reduce between-group differences with regards to gender ratio (75% vs. 73% males in controls vs. schizophrenia patients), age distribution (32.3 ± 9.8 vs. 33.4 ± 9.5

[2] http://schizconnect.org
[3] https://openfmri.org

years old) and motion levels (average frame displacement = 0.15 ± 0.05 vs. 0.17 ± 0.06, see Data preprocessing). Matching of schizophrenia and control subjects was achieved based on propensity scores, using the Optmatch R library version 0.9-7[4]. The propensity score associated with each participant was defined by the conditional probability of being in the clinical or control group given the confounding covariates (gender, age and motion). Propensity scores were then used to balance those covariates in the two groups. Although we took great care in matching participants with respect to these factors of no interest, it is very likely that other confounds such as medication in schizophrenia patients impacted the reported findings.

3.3.3 Data preprocessing

Brain imaging data preprocessing and extraction of functional brain connectomes were performed with the NeuroImaging Analysis Kit version 0.12.17 (NIAK[5]). Briefly, preprocessing included slice timing correction, estimation of rigid-body motion within the functional runs, nonlinear coregistration of the structural scan in stereotaxic space, individual coregistration between structural and functional scans, resampling of the functional scans at 3mm isotropic resolution in stereotaxic space, scrubbing of volumes with excessive motion (frame displacement greater > 0.5 mm), regression of confounds (slow time drifts, average of conservative white matter and cerebrospinal fluid masks and motion parameters), and smoothing of functional volumes with a 6 mm isotropic Gaussian blurring kernel. A detailed description of the preprocessing pipeline can be found at [6].

Individual functional connectomes included 2016 functional connections between 64 brain parcels. The functional brain parcellation was previously obtained by conducting a bootstrap analysis of stable clusters (BASC, Bellec et al. (2010b)) on an independent fMRI dataset of 200 healthy young subjects[7]. In each schizophrenia or control participant, the time series of a brain parcel consisted in the average of the voxel signals in the parcel. Connectivity measures between pairs of parcels were defined by Pearson

[4] https://cran.r-project.org/web/packages/optmatch/index.html
[5] http://niak.simexp-lab.org
[6] http://niak.simexp-lab.org/pipe_preprocessing.html
[7] https://doi.org/10.6084/m9.figshare.1285615.v1

product-moment correlation coefficients. Individual connectomes were parcel by parcel (64 x 64) symmetrical matrices that summarized connectivity levels in the whole brain. Lower triangular matrices were then vectorized for all subjects in order to form a subject by connections (382 x 2016) matrix.

3.3.4 Data analysis

Classification analyses were performed with a linear support vector machine (SVM) algorithm, as implemented in the SciKit-Learn python library version 0.18.1 (Abraham et al. 2014). The SVM classifier, a supervised classification algorithm, represented subjects as points in space, mapped so that the subjects of the separate clinical labels were divided by a clear gap (called a margin) that was as wide as possible. The hyperparameter C of the SVM was optimized using nested cross-validation. Each model used the residuals from a regression of confounding variables (gender, age and motion parameters) across connections estimated from the subjects selected for training the model. The evaluation metrics were computed using four main values, namely the number of true and false positive (TP, FP) as well as true and false negatives (TN, FN). Sensitivity was defined as TP/(TP+FN), specificity as TN/(TN+FP) and accuracy as (TP+TN)/(TP+FP+TP+FN). The main analyses evaluated the impact on classification accuracy of the number of site(s) (1, 2, 3, 4 or 5) included in the training set. We evaluated this impact in situations where the test set included only subjects from the same site(s) used during training (intra-site test with 10-fold cross validation) or, alternatively, situations where the test set included only subjects from sites not used during training (inter-site test with "leave-site-out" cross validation). Cross validation ensured that the subjects used for training were never used in the test phase.

The statistical significance of changes in accuracy levels as a function of the number of sites used for training and whether data used for testing were drawn from the same dataset(s) used for training (intra-site vs inter-site) was assessed with binary logistic regressions using the GLM function in R version 3.2.5. These analyses relied on the prediction of categorical outcomes (hit/miss data) based on predictor variables (number of sites used for training, intra-site vs inter-site). Significance threshold in the different

contrasts was set at $p < 0.05$.

Complementary analyses were conducted. First, we explored differences in whole brain connectivity between schizophrenia patients and controls using mass univariate statistics for the various training site combinations. Similarly for multivariate classification analyses, we extracted feature weights separately for all site combinations. We then examined the level of correspondence across site combinations for both univariate and multivariate analyses. Second, we aimed at demonstrating the presence of multivariate site effects on functional brain connectivity. To this end, we determined accuracy levels for the classification of scanning sites by performing separate SVM analyses for all pairs of sites, using 10-fold cross validation as in the main analyses.

3.4 Results

3.4.1 Correspondence across site combinations

We first report patterns of functional brain dysconnectivity in schizophrenia patients based on mass univariate statistics. For the sake of interpretability, the 64 brain parcels were sorted in relation to 7 large-scale brain networks from the same multi-scale functional brain atlas (Figure 1a,b). When pooling data from all subjects and sites, a connectome-wide association analysis revealed widespread decreased connectivity in schizophrenia patients (Figure 1c), with 769 out of 2016 connections exhibiting a significant effect after false discovery rate correction (qFDR < 0.05). Differences between schizophrenia patients and controls were further examined separately for each unique combination of 1 to 5 training sites (61 possibilities: 1, 2, 3, 4, 5, 6, 1-2, 1-3, 1-4, ..., 1-2-4-5-6, 1-3-4-5-6, 2-3-4-5-6). Results revealed high variability in the nature of mass univariate effects across training site combinations, with small correlation between them when there was no overlap between site combinations (Figure 1c). By contrast, large correlations were observed when site combinations overlapped. Weight matrices, which indicate for each connection the importance of that connection in the decision process, were also extracted for the whole sample as well as each site combination in multivariate classification analyses (Figure 1d). The correspondence between site com-

binations mimicked the patterns of correlations seen for univariate analyses, with a large correlation between weights for site combinations that overlapped but a small correlation otherwise.

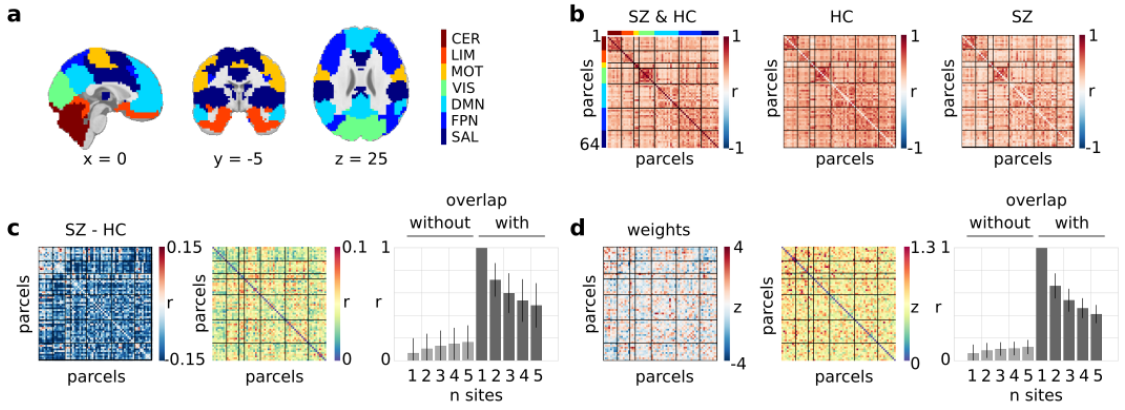

Figure 1: Sorting the 64 brain parcels in relation to 7 large-scale brain networks (a) allowed to best reveal the structure of whole-brain connectivity in both schizophrenia patients and controls, here shown when combining data from all sites (b). Results for univariate analyses (c). The left panel reports mean differences in connectivity between schizophrenia patients and controls pooled across all sites. The middle panel shows variations (standard deviation) of between-groups differences across the various site combinations. The right panel reports correlations of univariates effects of schizophrenia between single sites of reference (sites 1 to 6) and various site combinations as a function of whether site combinations included the sites of reference and the number of sites included in the combination (mean and standard deviation across the 6 reference sites). Results for multivariate classification analyses are similarly organized (d). The left panel provides the normalized weights obtained when pooling all subjects from all sites. The middle panel indicates how these weights vary across site combinations (standard deviation). The right panel provides correlations of weight matrices between single sites of reference and various site combinations. Abbreviations for networks are as follows: CER, cerebellum; VIS, visual; LIM, limbic, MOT, motor; SAL, salience; FPN, fronto-parietal; DMN, default-mode.

3.4.2 Classification findings

Classification of sites was performed with high accuracy (84%), indicating a significant multivariate impact of scanning site on functional brain connectivity. Training on

data from a single site led to a poor generalization of diagnosis classification to subjects drawn from another site, i.e. classification accuracy was much lower in the inter-site than intra-site configuration when only one site was used for training ($p < 0.005$). However, increasing the heterogeneity of the training set by including data from different sites improved accuracy of the classifier applied to another unknown scanning site and cognitive context ($p < 5x10^{-8}$) (Figure 2a). This compensatory effect was such that inter-site classification reached similar accuracy performance to intra-site classification when 5 different sites were used for classifier training ($p = 0.56$), thus suggesting excellent generalization in this context. The benefit of using heterogeneous training data when classifying subjects drawn from the same sites as the training set was much more moderate than for the inter-site training-test configuration, yet was significant ($p < 0.05$). Formal testing of an interaction effect revealed a significant effect ($p < 0.05$), thus demonstrating that improved generalization on novel sites following multisite training was not merely a consequence of increasing sample size.

3.5 Discussion

The present findings highlight a prerequisite for an optimal translation of classification tools from the research to clinical realm. Namely, classifier training should be performed on data that are sufficiently representative of sites and/or mental state variations in order to generalize well for large-scale clinical use. In particular, the accuracy scores reported in most of the existing literature should be interpreted with caution, as they only reflect within-site generalizability and may therefore overestimate the accuracy.

Mass univariate analyses evidenced brain dysconnectivity across the entire brain, with significant effects in over a third of brain connections distributed in various large-scale brain networks, from cognitive to primary sensory networks. Abnormally decreased rather than increased functional connectivity in schizophrenia is largely consistent with previous reports in the literature (Pettersson-Yeo et al. 2011). With close to 200 schizophrenia patients and 200 controls, our fMRI connectome-wide association

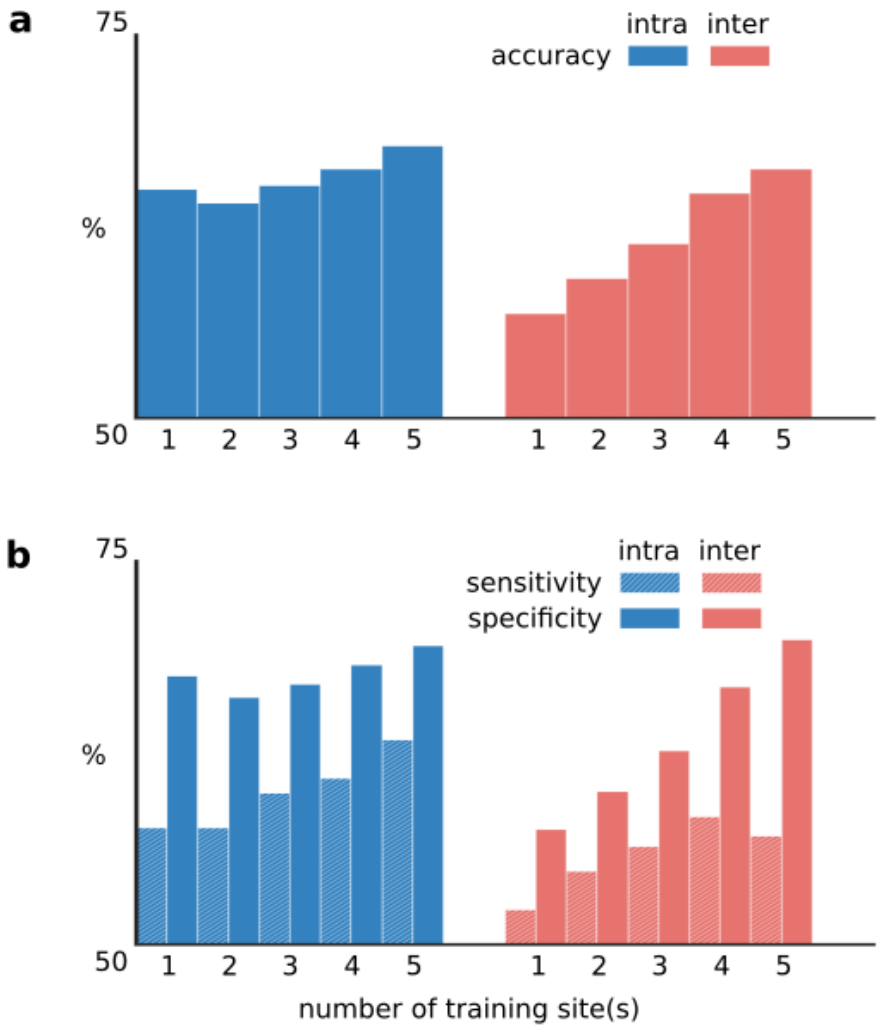

Figure 2: Histograms show the percentages of classification accuracy (a) as well as sensitivity and specificity (b) as a function of the number of sites (1, 2, 3, 4 or 5) from which individual connectomes were drawn for training, and whether testing was performed on subjects drawn from the same site(s) as during training (intra-site test) or not (inter-site test).

analysis is one of the largest to be reported to date. The present work and previous similar studies conducted in schizophrenia (Cheng et al. 2015c, Schilbach et al. 2016, Skåtun et al. 2016) underscore the utility of pooling data across multiple sites of acquisition in order to achieve higher sample size and more reliable findings. The marked variability in dysconnectivity patterns detected in each site separately could induce a deleterious effect of multisite data pooling on statistical power as compared to data obtained at a single site. However, there is support to claim that the initial deleterious effect of multisite data pooling can be mitigated by the increase in sample size and number of sites in the context of intra-site mass univariate as well as multivariate analysis (Dansereau et al. 2017). This is in line with our findings suggesting that a similar compensating effect can be obtained using multisite data aggregation in inter-site multivariate prediction, a configuration that is most likely to be found in clinical settings.

Multivariate classification analyses indicate that increasing sample size through multisite data pooling increased diagnosis prediction in schizophrenia, although this effect was of small amplitude. More critically, an additional benefit of including heterogeneous data was that the classifier generalized better to data that were not represented during training, neither in terms of scanning site nor mental content. This demonstration is concordant with a previous report that classified gender as a proof-of-concept application (Huf et al. 2014), and underscores the benefit of pooling multisite data for the purpose of generalizability and clinical use. The observed gain of almost 10% in classification accuracy is appreciable. It is nonetheless noteworthy that the highest accuracy of schizophrenia diagnosis classification was below 70%, which precludes the immediate translation of such machine learning tools in the clinical realm. Beyond the fact that most classification work has to date investigated within-site generalizability, it is notable that most studies relied on small samples. This is likely to be accompanied by a publication bias by which only the most significant findings were published. While the average classification of schizophrenia diagnosis over 30 published studies is above 80%, it was accordingly shown that studies with a large sample size in fact reported lower classification accuracy (Arbabshirani et al. 2017). Besides, low classification accuracy is very likely dependent on the ill-definition of clinical labels, as schizophrenia

is a highly heterogeneous psychiatric disorder (Kapur et al. 2012). The stratification of patients into more homogeneous neurobiological subtypes, beyond clinical symptoms, will likely define more precise labels that will lead to improved classification of their diagnosis. The characterization of mental illness heterogeneity through the identification of such different biotypes, in particular based on fMRI brain connectivity, is a topic of burgeoning research in various psychiatric disorders (Dias et al. 2015, Drysdale et al. 2017, Gates et al. 2014).

It is anticipated that neuroimaging-based classification will ultimately assist psychiatrists in not only diagnosis but also prognosis and theragnosis in mental illness, including schizophrenia. The future integration of classifiers into mental health care will require studies with dramatically increased sample size (Dubois and Adolphs, 2016). Most studies indeed suffer from insufficient data, possibly resulting in biased accuracy estimation, under-representation of mental illness heterogeneity and unstable findings (Arbabshirani et al. 2017). Future work will also need to develop novel algorithms with improved capabilities and to better define clinical labels. The present work identifies one specific parameter that will facilitate an optimal translation of supervised machine learning into clinical practice, namely the need to train classifiers on data that are sufficiently representative of heterogeneity with regards to scanning sites and mental contents.

3.6 Acknowledgments

Data from one study (emotional memory task) were collected thanks to grants from the Canadian Institutes of Health Research, Gender and Health Institute to Dr Mendrek (CIHR grant number 200603MOP-158161-GSH-CFCA-130656). Data from 4 other studies were accessed through the SchizConnect platform[8]. As such, the investigators within SchizConnect contributed to the design and implementation of SchizConnect and/or provided data but did not participate in analysis or writing of this report. Funding of the SchizConnect project was provided by NIMH cooperative agreement 1U01 MH097435. SchizConnect enabled access to the following data repository: the COllab-

[8] `http://schizconnect.org`

orative Informatics and Neuroimaging Suite Data Exchange tool (COINS[9]). Data from one study (resting-state) was collected at the Mind Research Network and funded by a Center of Biomedical Research Excellence (COBRE) grant 5P20RR021938/P20GM103472 from the NIH to Dr Vince Calhoun. Data from two studies (oddball and Sternberg item recognition paradigm tasks) were obtained from the Mind Clinical Imaging Consortium database. The MCIC project was supported by the Department of Energy under award number DE-FG02-08ER6458. MCIC is the result of efforts of co-investigators from University of Iowa, University of Minnesota, University of New Mexico and Massachusetts General Hospital. Data from a fourth study (N-back task) were obtained from the Neuromorphometry by Computer Algorithm Chicago (NMorphCH) dataset[10] As such, the investigators within NMorphCH contributed to the design and implementation of NMorphCH and/or provided data but did not participate in analysis or writing of this report. The NMorphCH project was funded by NIMH grant RO1 MH056584. The last study (task-switching) was obtained through the OpenFMRI project[11] from the Consortium for Neuropsychiatric Phenomics (CNP), which was supported by NIH Roadmap for Medical Research grants UL1-DE019580, RL1MH083268, RL1MH083269, RL1DA024853, RL1MH083270, RL1LM009833, PL1MH083271, and PL1NS062410. Data analysis was supported by a grant from the Natural Sciences and Engineering Research Council of Canada (NSERC grant number #436141) to PB. CD is supported by a bursary from the Lemaire foundation.

[9] http://coins.mrn.org/dx

[10] http://nunda.northwestern.edu/nunda/data/projects/NMorphCH

[11] http://openfmri.org

CHAPTER 4

A BRAIN SIGNATURE HIGHLY PREDICTIVE OF FUTURE PROGRESSION TO ALZHEIMER'S DEMENTIA

In preparation.

C. Dansereau, A. Tam, A. Badhwar, S. Urchs, P. Orban, P. Rosa-Neto, P. Bellec

4.1 Abstract

Alzheimer's disease develops slowly over years or even decades before the apparition of clinical symptoms and, eventually, dementia. Many works have aimed at finding biomarkers able to accurately predict future progression to dementia in individuals with mild cognitive impairment. Unfortunately, patients diagnosed with Alzheimer's dementia represent a highly heterogeneous group from the standpoint of the brain pathophysiology. Accurate prediction of progression to dementia more than one year before onset has therefore proved very challenging. In this work, we propose a new machine learning technique that identifies a subgroup of patients for which clinical predictions can be made with high precision, i.e. the vast majority of selected individuals will eventually progress to dementia. We demonstrate here that it is indeed possible to train a model to reach high specificity (97%) and precision (90%) when predicting future progression to dementia up to three years before onset, in the cohort assembled by the Alzheimer's disease neuroimaging initiative (ADNI). The model only achieved moderate sensitivity (47%) because it was designed to target a specific brain signature mixing spatial patterns of atrophy and functional dysconnectivity. This multimodal, highly predictive brain signature was extracted from magnetic resonance images only, yet it was systematically accompanied by deposition of beta amyloid plaques, a hallmark of Alzheimer's disease. Our results represent a marked improvement on the current state of the art in terms

of specificity (about 10% increase) and precision (about 16% increase). The signature was in addition identified on patients suffering from dementia first, demonstrating that typical patterns of neurodegeneration are already present in prodromal individuals. Our approach provides a feasible method to select individuals with very high risk of progressing to dementia several years before the onset. We believe this technology has a lot of potential to enrich the recruitment in clinical trials, and help demonstrate, or invalidate, the efficacy of new interventions. The method we used is relatively simple, building on well established machine learning tools, and may prove useful in the future for a wide range of other applications where the targets for predictions are noisy or heterogeneous, as is often the case in medicine.

4.2 Introduction

Alzheimer's disease (AD) is the most common age-related neurodegenerative disorder. The typical progression of late-onset, sporadic AD comprises a lengthy preclinical stage, a prodromal stage of mild cognitive impairment (MCI), and a final stage of dementia. Usually, by the time patients experience the dementia phase, severe and irreversible neurodegeneration has already occurred. In order to be effective, therapies should likely be initiated at earlier stages of the disease, when some markers of the disease are apparent but the symptoms have not yet appeared. For this reason, many works have aimed at finding biomarkers that can predict future progression to AD dementia at the prodromal or even preclinical stages (Orban et al. 2017c, Rathore et al. 2017). Accurate prediction beyond two years has however proven to be challenging, likely due to the considerable pathophysiological heterogeneity underlying existing clinical diagnosis (Rathore et al. 2017). We propose here to address the heterogeneity issue by identifying a subset of individuals with MCI who share a common brain signature predictive of oncoming AD dementia with high precision.

A clinical diagnosis of Alzheimer's dementia is primarily established on the basis of symptoms. To qualify for dementia, these symptoms need to interfere with a patient's ability to function in daily activities. Dementia is considered to be probably due to AD

when the symptoms appear gradually and the most prominent deficits fall either into an amnestic (i.e. memory, the most common) or nonamnestic category, i.e. language, visual or executive. There also needs to be no dominant symptoms suggestive of another type of neuropathology such as Lewy bodies, fronto-temporal atrophy or vascular abnormalities (McKhann et al. 2011).

The actual cause of dementia, AD or otherwise, can currently only be confirmed by a post mortem pathophysiological examination. The hallmarks of AD are the accumulation of beta-amyloid plaques and tau protein neurofibrillary tangles in the brain, as well as marked atrophy of the medial temporal lobe. The analysis of Beach et al. (2012) revealed an important mismatch between clinical and histopathological diagnoses: sensitivity ranged from 71% to 87% and specificity ranged from 44% to 71%, depending on the level of confidence in the clinical and pathophysiological examination. In particular, 30% of patients diagnosed with AD dementia in that study had no or very minimal signs of AD pathology in their brains. In addition to such incorrect diagnoses, comorbidity of neurodegenerative diseases was highly prevalent, that is a co-occurrence of two or more disorders including AD, cerebrovascular disease, Lewy body disease, or frontotemporal degeneration Jellinger et al. (2014), Rabinovici et al. (2017). Biomarkers of AD can also be observed in 10% to 30% of cognitively normal (CN) individuals, as well as 40% of patients diagnosed with non-AD dementia (Beach et al. 2012). Finally, plaques and tangles are general markers of brain injury that are not unique to AD, e.g. they are seen in patients with brain traumatic injuries (Marklund et al. 2009). Distinct pathways are likely involved in AD, and subtypes of AD pathophysiology may emerge in the future (Au et al. 2015). In summary, the clinical labels of neurodegeneration currently used are often incorrect (wrong underlying disease), incomplete (missing several interacting diseases) or unspecific (pooling together different pathways with overlapping biomarkers).

To better diagnose AD in vivo, many imaging techniques have been developed to track the propagation of key markers, both across brain regions and over time. Both beta-amyloid and tau can be imaged in vivo using Positron Emission Tomography (PET) (Fodero-Tavoletti et al. 2011, Sperling et al. 2011). Structural magnetic resonance imaging (MRI) provides a non-invasive measure of temporal lobe atrophy, as well as de-

creases in cortical thickness throughout the brain (Lerch et al. 2005). Large imaging samples such as the one collected by the Alzheimer's Disease Neuroimaging Initiative (ADNI) have established that beta-amyloid and tau starts accumulating years, possibly decades, before the onset of clinical symptoms, and that atrophy also typically precedes the onset of clinical symptoms (McConathy and Sheline 2015). Imaging biomarkers are increasingly used to complement neuropsychological testing to diagnose AD (Dubois et al. 2007). In recent years, a great amount of work has been devoted to the identification of novel or more sensitive imaging based biomarkers of AD and MCI using machine learning techniques (Rathore et al. 2017). The current state of the art on predictive models using the ADNI dataset reached 95% accuracy (precision of 96%, specificity of 95% and sensitivity of 92%) to classify AD vs cognitively normal (CN) (Fan et al. 2008b, Xu et al. 2015, Zhu et al. 2014, Zu et al. 2016) and 80% accuracy (precision of 80%, specificity of 75% and sensitivity 85%) to identify patients with MCI who will progress to AD dementia in the next three years (Cheng et al. 2015a;b, Korolev et al. 2016, Moradi et al. 2015, Toussaint et al. 2012, Zheng et al. 2015), using mainly anatomical, FDG-PET or amyloid-PET measures. The prediction accuracy for progression to dementia however plummeted after 1.5 years, reaching only 75% accuracy over a 18-36 months time window (Arbabshirani et al. 2017, Korolev et al. 2016).

These accuracy scores however need to be properly interpreted. Korolev et al. (2016), in particular, took great care of separately reporting the specificity (76%, proportion of stable MCI being correctly classified), sensitivity (83%, proportion of progressor MCI being correctly classified), and precision (80%), i.e. the proportion of actual progressors amongst individuals classified as such. Precision is a key metric for enrichment in a clinical trial, as it dictates how many patients will decline in the absence of treatment. For a given sensitivity and specificity, the precision does depend on the baseline ratio between stable and progressor MCI in the sample. Working on a scenario of 30% MCI progressors in the cohort (which matches actual rates seen in clinical populations), Korolev et al. (2016) would have an expected precision of 60.2%. There is therefore ample margin for improvements in terms of prognostic precision of future progression to AD dementia within 3 years. We note that, because there are more MCI stables than progressors at

baseline, an increase in specificity has higher leverage on precision than sensitivity. For example, working again on a scenario of 30% MCI progressors in the cohort, with an equal sensitivity and specificity at 80%, the precision is 63%. An increase of 10% sensitivity will only increase precision by 3%, while an increase of 10% specificity will boost precision by 14%.

We hypothesized that the precision of imaging-based diagnosis of AD in past studies was severely limited by the pathophysiological heterogeneity of clinical cohorts. In this work, we proposed a new machine learning technique that aims at identifying a subgroup of patients for whom clinical predictions can be made with high precision. We specifically trained a model to sacrifice sensitivity for high specificity and precision (Figure 1A). The behaviour of the proposed method is illustrated by a simple simulation (Figure 1B). The task was to classify two classes using a separation line, represented by blue dots for controls and red dots for patients. The distribution of both red and blue subjects was heterogeneous, in the sense that each distribution was a mixture of several Gaussian classes. Some of these classes were clearly separable, yet others were not, with blue and red points closely overlapping (maybe because of incorrect, incomplete or unspecific diagnoses). When a standard classifier is applied on that data, it identifies a separation line making a tradeoff in sensitivity and specificity across all examples (see Figure 1B, second column). By perturbing the data, it is possible to identify the "easy cases", i.e. the data point that can be correctly and reliably classified: more opaque points are associated with more reliable predictions and clearly identify the two well-separated classes at the top in Figure 1B, third column. A separate model is then trained to identify the "easy cases" red points (see Figure 1B, fourth columns). The resulting prediction of red labels has limited sensitivity, as the problematic cases are not being detected at all, but it has near perfect specificity and precision. Note that the example found in Figure 1B was actually computed with our proposed method. To evaluate how this method managed to extract AD biomarkers, we first examined the classic problem of predicting clinical diagnosis in a cohort including CN participants and patients with AD dementia. We then used the model trained on the CN vs AD classification problem to make predictions on the subjects with MCI. Our hypothesis was that a brain signature of

AD dementia would already be present at the prodromal stage, and predictive of future progression to dementia. To test this hypothesis, we evaluated whether the MCI patients flagged as AD would develop AD dementia within three years.

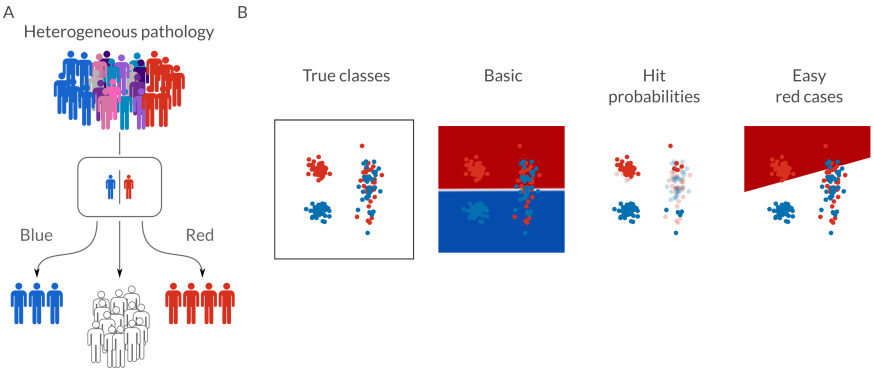

Figure 1: Panel A show the identification of easy cases for each class, Panel B prediction of clinical labels in a two-class problem, in the presence of heterogeneous labels in a subset of the data. First column show the initial classification problem with the distribution of the two classes. The second column show a basic classifier decision hyperplane. Third column show the subject that have been flagged as high hit probability in hard color and the low hit probabilities with some transparency. Fourth column show the final decision hyperplane of the red subject with the HPS signature.

4.3 Results

Multimodal imaging markers

We extracted multimodal measures of brain organization, that could be used for automated AD diagnosis. The measures were derived from the baseline MRI scans of the Alzheimer's Disease Neuroimaging Initiative 2 (ADNI2) cohort, which included anatomo-functional imaging for CN subjects (N=49) as well as patients suffering from AD dementia (N=24) (available sample size after quality control on 10/2016). We decided to include a range of different measures as a basis for diagnosis, which have previously been shown to be sensitive markers of AD dementia. These included gray matter (GM) thickness (Eskildsen et al. 2013, Querbes et al. 2009), GM volume of various brain structures (Karas et al. 2004), as well as seed-based fMRI connectivity maps generated

for 20 intrinsic connectivity brain networks (Bellec et al. 2015).

Substantial inter-individual variations were observed in the brain distribution of imaging measures. For example, some subjects showed higher- or lower-than average volumetric measures across extensive brain territories, such as the right medial occipital cortex in subject 1 (lower) and subject 73 (higher), see Figure 2A. We investigated whether such patterns could be found systematically in a subgroup of subjects. For this purpose, we quantified the similarity of GM volume maps between any given pair of subjects using a Pearson correlation coefficient (Figure 2B). A cluster analysis revealed the presence of three subgroups of subjects with homogeneous GM volume maps. These subgroups were apparent as diagonal squares on the inter-subject similarity matrix with high similarity values, Figure 2B. These squares outline all subject-to-subject similarities within a specific subgroup. By contrast, low similarity values were observed in elements outside of these squares, which corresponded to pairs of subjects falling into different subgroups. A subtype template was generated for each subgroup by averaging maps of individuals within that subgroup, Figure 2B). In particular, subtypes 2 and 3 of GM volumetric maps reproduced the pattern in the occipital cortex observed in subjects 1 and 73, respectively. The separation between clusters was not clear-cut in matrix 2B, suggesting a continuum rather than discrete subtypes. We thus extracted a continuous measure (Pearson's correlation) of similarity, called subtype weight, between each individual map and each subtype map, Figure 2D). The subtyping procedure outlined above was applied independently for each type of measure (volumetric, cortical thickness, rs-fMRI) and each brain network (for rs-fMRI). We concluded by visual inspection to the presence of at least three subtypes for each modality/network, which we thus selected as a common number of subtypes across all modalities/networks for subsequent analyses.

Prediction of AD

We established a baseline performance for automatic classification of CN vs AD subjects using a well established machine learning model, i.e. a linear support vector machine model (SVM) (Cortes and Vapnik 1995). The model reached 70% precision (specificity 86%, sensitivity 67%) using tenfold cross-validation and multimodal (fMRI

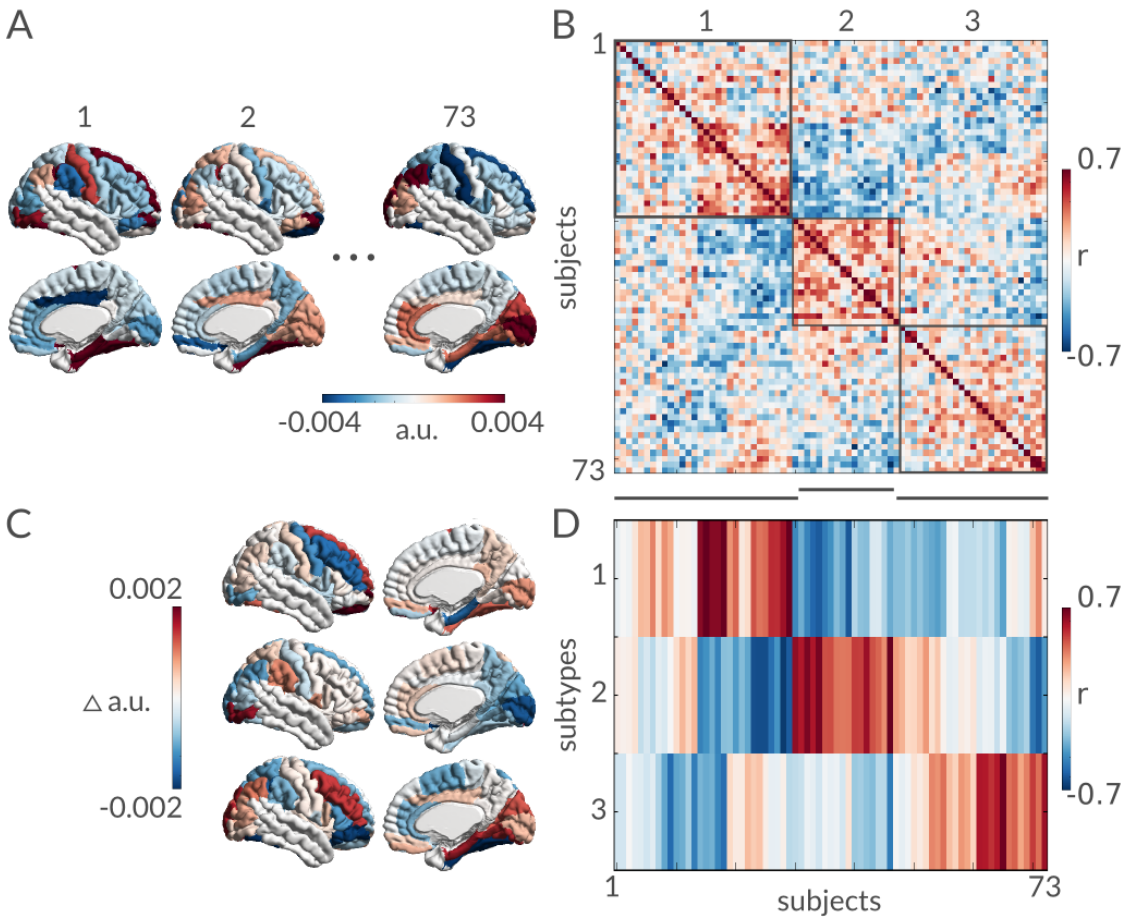

Figure 2: Demeaned gray matter volume measures of the right hemisphere. Panel A shows individual maps and the correlation of every subject with all other subjects in Panel B. Panel C shows the subtypes templates representing subgroups in the dataset. Panel D shows the association of each individual map in A with the each subtype template in C.

+ sMRI) subtype weights, Figure 3. The performance of the method trained on only fMRI 38% precision (specificity 47% and sensitivity 67%) and sMRI data only had a very close performance, with 67% precision (specificity 84%, sensitivity 67%). Note that, during cross-validation, the training of the model included both the generation of subtypes and the optimization of the SVM parameters.

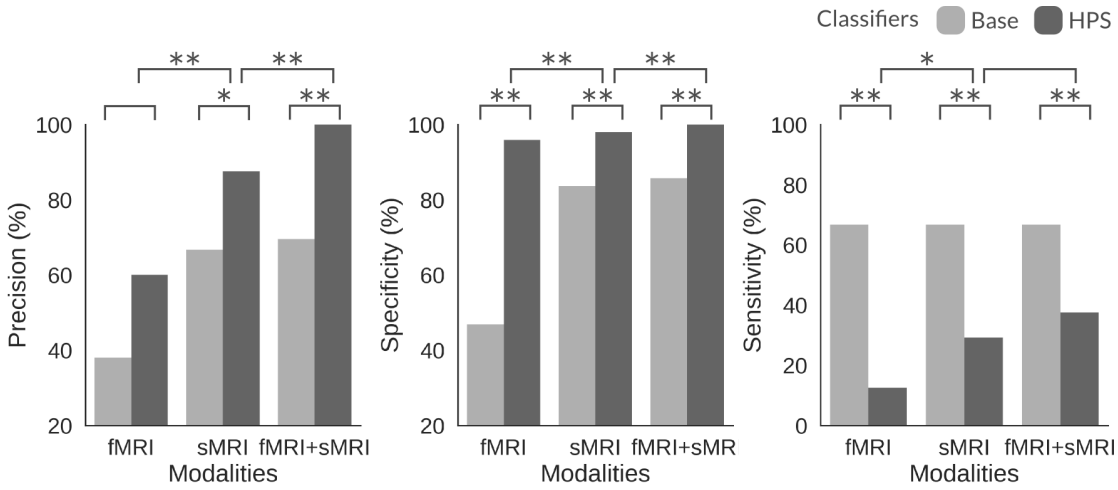

Figure 3: Figure shows the precision, specificity and sensitivity of the three modalities (fMRI, sMRI and fMRI+sMRI) at each stage (Base: basic classifier and HPS: highly predictive signature). Significant differences are shown with $*$ for $p < 0.05$ and $**$ for $p < 0.001$).

Identifying easy cases

As we outlined in the introduction, the core idea of our approach was to identify a subset of subjects for which clinical labels are easy to predict, such as the points on the left in Figure 1A. To identify these "easy cases", we randomly perturbed the input data of the SVM model many times through subsampling, and assessed the hit probability for any given subject to be properly classified. We found that 68% of individuals had a perfect (100%) hit probability, with a small subset of subjects (18%) exhibiting less reliable predictions (hit-probability $< 90\%$), Supplementary material S1). We defined the "easy cases" as the subgroup of individuals reaching perfect hit probability.

Predicting easy cases

The next step of the method was to train a logistic regression (Fan et al. 2008a) to predict the AD "easy cases" Figure 6B, analogous to the rightmost column of Figure 1B. The full multi-stage process of subtype extraction, hit probability estimation and logistic regression was cross-validated using a ten-fold scheme in order to generate the

performance of the prediction of AD "easy cases". A perfect 100% precision (specificity 100%, sensitivity 36%) was reached for AD "easy cases", using multimodal structural and functional features. A significant improvement (in precision and specificity $p < 0.001$) of the HPS compared to the performance of the method trained on only fMRI data reached a precision of 60% (specificity 96%, sensitivity 13%) and sMRI data reached a precision of 88% (specificity 98%, sensitivity 29%), see see Figure 3. Compared to the reference SVM model, with multimodal features, the precision of our proposed HPS model was improved by a wide margin (30%, $p < 0.001$), as well as the specificity (15%, $p < 0.001$), at the cost of a marked loss in sensitivity (30%, $p < 0.001$). See Supplementary material Table S2 for a list of the performance of each model.

Highly predictive brain signature

The logistic regression model used to predict AD easy cases is based on a set of coefficients, which give more or less weight to a particular subtype and modality. As such, the individuals flagged as AD easy cases can be seen as sharing a brain HPS, composed of combination of subtype maps. The logistic model may in theory ignore a subtype or an entire modality, by setting the corresponding weights to zero. In practice, we found that the HPS relied on all three types of measures (functional connectivity, cortical thickness, and gray matter volume), Figure 4A. To rank the contribution of each modality in the decision process, we computed the absolute sum of the coefficients for each measure, relative to the sum of all absolute coefficients (Figure 4B). The thickness was the most important measure (60%), followed by the volumetric measures (29%), and finally functional connectivity (11%). The highest contributions came from four subtypes of thickness: bilateral patterns of cortical atrophy in temporal, sagittal and frontal areas (one subtype per hemisphere), and bilateral, opposite patterns of increased thickness (one subtype per hemisphere), Figure 4C. Two lateralized volumetric subtypes showed gray matter volume loss in the left motor, and right frontal areas as well as a gray volume increase in the left frontal and limbic regions. Finally, one functional subtype was very noisy and barely contributed to the model, while the other highlighted a connectivity subtype connecting the visual network with frontal areas.

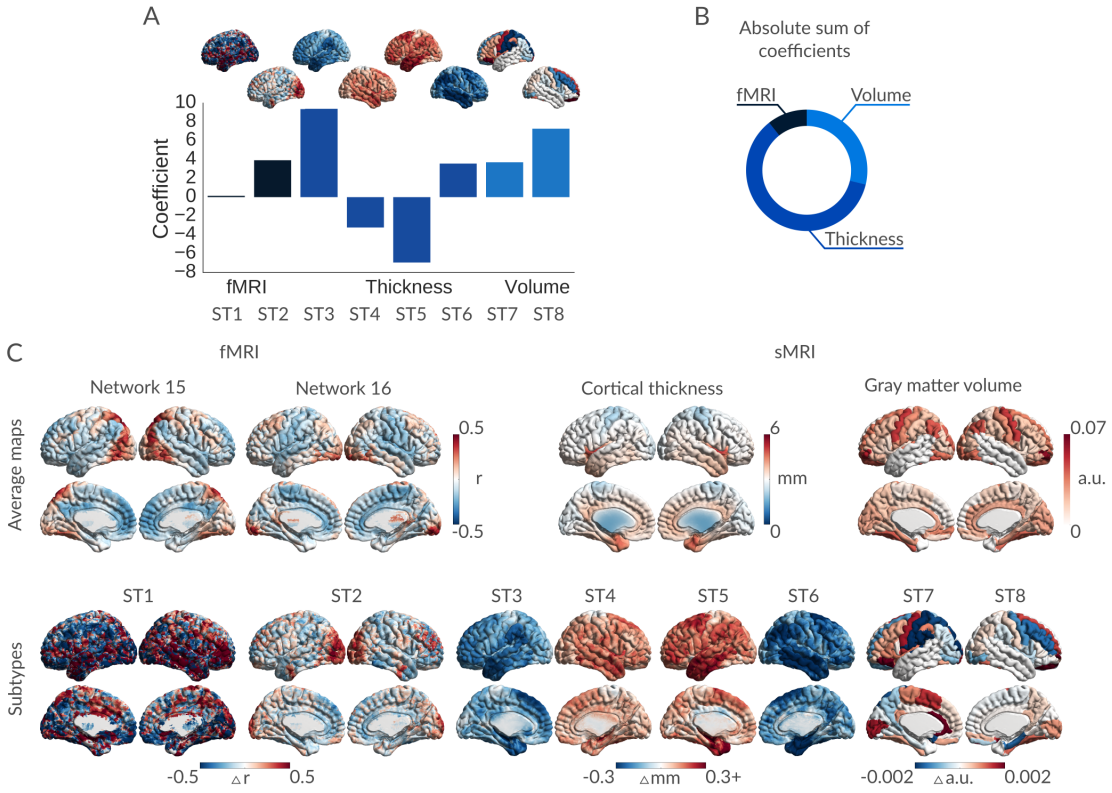

Figure 4: Panel A shows the contribution of each modality to the decision, the ratios are computed by the sum of the absolute coefficient for each modality. Panel B shows the coefficients of the high-confidence prediction model for each subtype map. Panel C shows, on top, the average maps for each modality and on the bottom the subtype maps used for the high-confidence prediction.

Prediction of progression to dementia

We applied the HPS model to patients with MCI from the ADNI2 cohort, with the hypothesis that those would likely progress to AD dementia. The imaging sample for this experiment included the baseline structural and functional scans of all patients with MCI in the ADNI2 cohort ($N = 79$). We further stratified the patients with MCI into stable MCI (sMCI, $N = 37$), i.e. most recent clinical status is MCI with at least 36 months follow up, and progressors (pMCI, $N = 19$), i.e. individuals whose most recent known clinical status is AD dementia, with progression from MCI to AD dementia occurring within 37 months. The HPS model selected a subset of 10 subjects. Using the longitudinal follow-up clinical data provided by ADNI2, we found that 9 out of 10 of these

subjects were pMCI (precision of 90%, specificity of 97%, sensitivity of 47%), compared to 34% pMCI in the whole MCI sample ($p < 0.001$), Figure 5A. Within the HPS subgroup, the time to progression from baseline to the first evaluation of AD dementia appeared uniformly distributed from 5 to 37 months, with 50% subjects progressing after 24 months (Figure 5C). In addition, 100% of the MCI participants flagged as HPS were positive for beta amyloid deposition with AV45 testing, compared to a 69% rate in the whole MCI sample ($p < 0.05$), Figure 5A. The rate of ApoE4 carriers in the HPS subsample was 78%, compared to 55% in the whole MCI group ($p > 0.05$), Figure 5A.

4.4 Discussion

The main goal of this work was to develop an imaging-based AD diagnosis with high precision and specificity. The proposed HPS approach did reach excellent performance in these respects, with 100% precision and specificity when distinguishing patients with AD dementia from CN participants and 90% precision, 98% specificity when predicting which MCI patients would progress to dementia, up to three years before onset (see Table S2). These results represent a sizable and significant improvement in precision and specificity over the state of the art on this task, see Table 1. No data from MCI patients were used to train the model, which removes the possibility of a bias due to improper cross validation. The only HPS subject with MCI improperly classified as a progressor did have a series of 4 notes attached to his visits in the ADNI database, reporting on a decline in cognitive performance at each visit, and a marked decline at the last visit. This decline was still not severe enough for a diagnosis of AD dementia. The subject had no follow up available after 36 months, for unknown reason.

The high specificity of the HPS model came at the cost of a limited sensitivity: 38% when distinguishing patients with AD dementia from CN participants, and 47% when predicting which MCI patients would progress to dementia, which is significantly less than most recent published models, see Table 1. The HPS model is not designed to be sensitive, as it is trained to recognize a particular, homogenous brain signature present in only a fraction of the participants. The results of (Beach et al. 2012) suggest that

Table 1: Supervised classification of MCI progression to AD dementia using the ADNI database. Progression time was establish if the the subject progresses to AD status in the next 36 months. Significant improvement of our method compared to each paper for the adjusted precision (adjusted for a pMCI ratio of 34% comparable to our sample) and specificity are shown with $*$ for $p < 0.05$ and $**$ for $p < 0.001$) and conversely significant decrease in sensitivity of our method compared to each paper.

Author	N(sMCI, pMCI)	Precision	Precision (adjusted)	Specificity	Sensitivity
Dansereau et al. (This paper)	37, 19	**90%**	**90%**	**97%**	47%
Mathotaarachchi et al. (2017)	230, 43	51%	74%	87%*	71%*
Korolev et al. (2016)	120, 139	80%	65%*	76%**	83%*
Moradi et al. (2015)	100, 164	85%	63%*	74%**	87%**
Eskildsen et al. (2013)	134, 149	70%	52%**	68%**	66%
Wee et al. (2013)	111, 89	77%	68%*	84%**	64%
Gaser et al. (2013)	62, 133	90%	70%*	84%**	71%*
Davatzikos et al. (2011)	170, 69	57%	63%*	71%**	**95%****
Koikkalainen et al. (2011)	215, 154	66%	58%*	71%**	77%*
Misra et al. (2009)	76, 27	42%	51%**	60%**	80%*

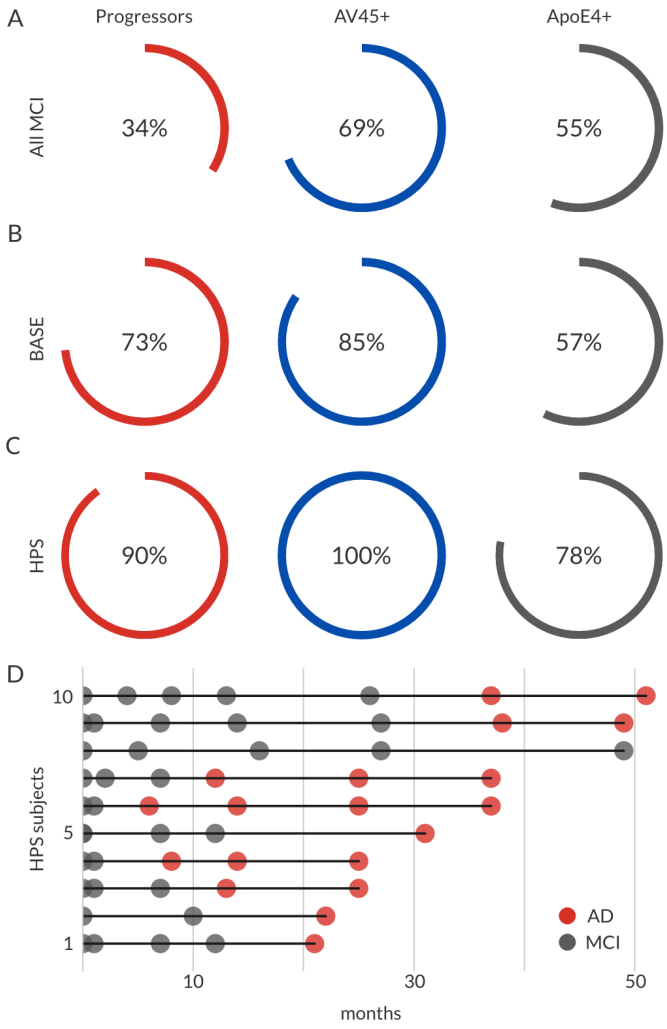

Figure 5: Statistic on the MCI showing the signature. Panel A shows the percentage of MCI who progress to AD, the percentage of subjects positive for beta amyloid deposits using the AV45 marker and the percentage of carriers of one or two copies of the ApoE4 allele for the entire MCI cohort. Panel B shows the same statistics for the selection of the base classifier while Panel C displays statistics for subjects flagged as HPS. Panel D shows the clinical status of each HPS subject over time from the baseline scan.

only about half of patients diagnosed with AD dementia have clear AD brain markers post-mortem. The observed sensitivity of 38%- 47% is thus consistent with the idea that the HPS model is picking on a typical brain presentation of AD that is already present at the prodromal stage of the disease. Note that there was no need for patients with MCI to have as much absolute measures of atrophy as patients with AD dementia to be

recognized as HPS, as long as these patients presented with a similar spatial distribution of the atrophy, relative to other brain regions.

The anatomical features selected by the method were in line with recent subtyping works, e.g. (Hwang et al. 2015), showing predominant atrophy in the temporal lobe, as well as the temporo-parietal juncture, in particular. The functional maps were more difficult to interpret, and seemed more noisy. They still made a significant improvement in the performance of the HPS model. Because of the regularization in the logistic regression used to build the HPS model, features coming from different modalities did compete to be selected in the model. If redundant features existed, the ones with largest predictive power were selected by the classifier. This may explain why the selected functional subtypes did not involve the regions showing atrophy in the structural subtypes. We hypothesized that the HPS inferred from the AD vs CN prediction would also be useful to predict if a subject at the prodromal stage (MCI) would progress to dementia. Our results did validate this logic, but alternative strategies may be investigated in the future, e.g. training a model directly on the progressor vs stable MCI task.

A limitation of the present study was a moderate sample size, with $N = 56$ patients suffering from MCI. Although the ADNI is a large database, resting-state fMRI has only been added to the protocol in the latter stages of the study, ADNI GO and ADNI2. In addition, fMRI was only acquired on a third of participants, even after it was added to the protocol. Because of the early role of synaptic dysfunction in AD, and the potential ability of fMRI to capture such dysfunction, we wanted to build an anatomo-functional diagnostic tool. But this choice did limit the sample size of our study. In the future, we are planning to replicate our findings using structural measures only, so we can use the entirety of ADNI. Even with a larger sample size, another limitation of the ADNI dataset is that it does not reflect the diversity of cases observed in real-life clinical practice. Participants were in particular screened to exclude vascular dysfunction, which is a common comorbidity in AD. Resources with more inclusive enrollment criteria will become important in the future to better assess the generalizability of a biomarker-based AD diagnosis.

The most direct application of the HPS model is population enrichment for phar-

maceutical clinical trials in AD (Mathotaarachchi et al. 2017, Woo et al. 2017). By recruiting almost exclusively patients who would normally progress to AD dementia, such enrichment would increase the effect size of the drug while reducing the sample size needed to demonstrate efficacy and therefore would also reduce the cost of the trial. The HPS brain signature is not shared among all the AD dementia population (making it a subtype), but is common enough to represent a substantial portion of participants of interest (about a third of AD dementia subjects and half of MCI progressors). An alternative enrichment strategy, more geared towards generalizability, would be to only exclude subjects that will very likely not progress to AD dementia. The HPS method thus brings us closer to precision medicine by proposing a middle ground between traditional clinical cohorts and an entirely individual medicine.

In this manuscript, we focused exclusively on two MRI modalities our rationale was that MRI is non-invasive and already widely used in patient care in elderly populations. Beta amyloid and tau PET imaging, by contrast, are more expensive and less available, while lumbar punctures are invasive. Nevertheless, it will be important in the future to see if a combination of PET imaging, blood tests looking for specific inflammatory proteins, cognitive scores, genetic factors, lifestyle factors, or others can help create multiple HPS that would in effect increase the sensitivity of the model for the early detection of Alzheimer's pathology.

4.5 Acknowledgments

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Alzheimer's Association; Alzheimer's Drug Discovery Foundation; BioClinica, Inc.; Biogen Idec Inc.; Bristol-Myers Squibb Company; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.;

GE Healthcare; Innogenetics, N.V.; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Medpace, Inc.; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Synarc Inc.; and Takeda Pharmaceutical Company. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health [1]. The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California. This research was also supported by NIH grants P30 AG010129 and K01 AG030514.

The computational resources used to perform the data analysis were provided by Compute Canada[2] and CLUMEQ[3], which is funded in part by NSERC (MRS), FQRNT, and McGill University. This project was funded by NSERC grant number RN000028 and the Canadian Consortium on Neurodegeneration in Aging (CCNA), through a grant from the Canadian Institute of Health Research and funding from several partners including SANOFI-ADVENTIS R&D. CD is supported by a salary award from the Lemaire foundation. PB is supported by a salary award from "Fonds de recherche du Québec – Santé" and the Courtois Foundation.

4.6 Materials and methods

Dataset

All functional and structural data were obtained from the Alzheimer's Disease Neuroimaging Initiative 2 (ADNI2) sample, a longitudinal standardized acquisition including three populations: cognitively normal subjects, patients with mild cognitive impairment and patients with dementia due to AD. All participants gave their written informed

[1] www.fnih.org
[2] https://computecanada.org/
[3] http://www.clumeq.mcgill.ca/

consent to participate in the ADNI2 study, which was approved by the local ethics committee of participating institutions across North America. The consent form included data sharing with collaborators as well as secondary analysis. The present secondary analysis of the ADNI2 sample was approved by the local ethics committee at CRIUGM, University of Montreal, QC, Canada. All resting-state fMRI and structural scans were acquired on 3T Philips scanners with 8 channels. We performed analyses on the first usable scan (typically the baseline scan) from ADNI2.

The acquisition parameters were as follows: structural scan 170 slices, voxel size 1x1x1.2 mm3, matrix size 256x256, FOV 256 mm2, TR 6.8 s, TE 3.09 ms, FA 9 degrees. A functional scan of 7 min, 48 slices, voxel size 3.3x3.3x3.3 mm3, matrix size 64x64, FOV 212 mm2, TR 3 s, TE 30 ms, FA 80 degrees, No. volumes 140. For detailed information on the acquisition, see www.adni-info.org.

Extraction of functional features

Each fMRI dataset was corrected for slice timing; a rigid-body motion was then estimated for each time frame, both within and between runs, as well as between one fMRI run and the T1 scan for each subject (Collins et al. 1994). The T1 scan was itself non-linearly co-registered to the Montreal Neurological Institute (MNI) ICBM152 stereotaxic symmetric template (Fonov et al. 2011), using the CIVET pipeline (Ad-Dab'bagh et al. 2006a). 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 frames showing a frame displacement, as defined in Power et al. (2012), greater than 0.5 mm were removed. An average residual frame displacement was also estimated after scrubbing for further group analyses. A minimum of 50 unscrubbed volumes per run was required for further analysis (13 subjects were rejected). The following nuisance covariates were regressed out from fMRI time series: slow time drifts (basis of discrete cosines with a 0.01 Hz highpass 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, Lund et al. 2006). The fMRI volumes were finally spatially smoothed with a 6 mm isotropic Gaussian blurring kernel. Datasets were preprocessed and analyzed using the NeuroImaging Analysis Kit - NIAK - version 0.12.17 (http://niak.simexp-lab.org), under CentOS with Octave (http://gnu.octave.org) version 3.6.1 and the MINC toolkit (http://bic-mni.github.io/) version 0.3.18. Preprocessing of MRI data was executed in parallel on the Guillimin supercomputer (http://www.calculquebec.ca/en/resources/compute-servers/guillimin), using the pipeline system for Octave and Matlab - PSOM (Bellec et al. 2012). Seed-based fMRI connectivity maps were obtained using a functional brain template of 20 networks covering the entire brain. The Pearson's correlation between the average time series of each network and every voxel of the brain was computed to derive one functional connectivity map per network.

Extraction of structural features

Native individual T1-weighted MRI scans were corrected for non-uniformity artifacts with the N3 algorithm (Sled et al. 1998). The corrected volumes were then masked for brain tissues (Smith 2002) and registered into stereotaxic space (Collins et al. 1994). The registered, corrected images were segmented into gray matter (GM), white matter (WM), cerebrospinal fluid (CSF) and background using a neural net classifier (Tohka et al. 2004). The WM and GM surfaces were extracted using the Constrained Laplacian-based Automated Segmentation with Proximities algorithm (Kim et al. 2005, MacDonald et al. 2000) and were resampled to a stereotaxic surface template to provide vertex based measures and lobar segmentation (Lyttelton et al. 2007). Cortical thickness was measured in native space using the linked distance between the two surfaces across 81,924 vertices (Im et al. 2008). Surface-based cortical thickness, as well as regional volume measures, were obtained using the structural MRI images processed using the CIVET 1.1.12 pipeline for each hemisphere as described in Ad-Dab'bagh et al. (2006b). The AAL template was applied on each hemisphere (40 regions per hemisphere) to extract the regional volumetric measures. The processing pipeline was executed on the Canadian Brain Imaging Network (CBRAIN) platform, a network of five imaging centers and eight High-Performance Computers for collaborative sharing and distributed

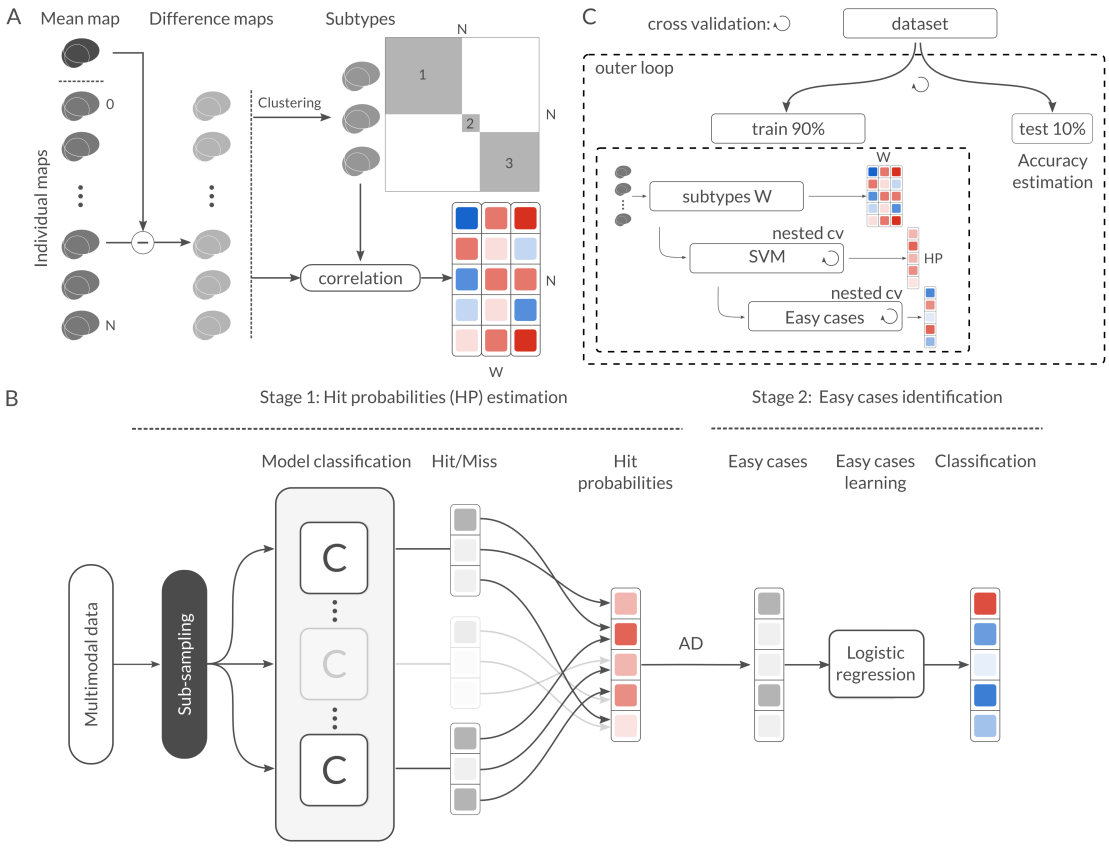

Figure 6: Panel A shows the feature extraction method called subtypes weights, Panel B framework workflow: stage 1 shows the hit probability computation based on random sub-sampling and stage 2 shows the training of dedicated classifier for each "high-confidence" signature. Panel C shows the nested cross-validation scheme used in this method.

processing of large MRI databases (Frisoni et al. 2011).

Multimodal imaging subtypes

We extracted subtypes that characterize the interindividual variability within the sample comprising CN and AD participants (at the time of scanning), independently for each type of measure (functional maps, cortical thickness maps, and volumetric maps). In order to reduce the impact of some factors of no interest that may influence the clustering procedure, we regressed out the age, sex, and average post-scrubbing frame displacement from individual maps, using a mass univariate linear regression model at each voxel. For

each type of brain measure, we derived a spatial Pearson's correlation coefficient between all pairs of individual maps. This defined a subject x subject similarity matrix (of size 73 x 73), which was entered into a Ward hierarchical clustering procedure, as implemented in SciPy version 0.18.1 (Jones et al. 2001–, Walt et al. 2011). We arbitrarily selected three subgroups for each type of measure, based on a visual examination of the similarity matrix. For each type of measure, the average map of each subgroup defined a subtype. For each individual, we computed the spatial correlation of their map with each subtype. The resulting weight measures formed a matrix of size (number of subjects) x (number of subtypes), which was used as the feature space in the predictive models used throughout the rest of the methods. Note that this entire subtyping procedure, including regression of confounds, was latter entered in a cross-validation scheme to assess the performance of the predictive models.

Prediction of AD

The baseline prediction accuracy was obtained by training a SVM model with a linear kernel, as implemented in Scikit-learn Pedregosa et al. (2011) version 0.18. A tenfold cross-validation loop was used to estimate the performance of the trained model. Classes were balanced inversely, proportional to class frequencies in the input data for the training. A nested cross-validation loop was used (stratified shuffle split (50 splits, 20% test size)) for the grid search of the hyper-parameter C (grid was 10^{-2} to 10^{1} with 15 steps). Note that the C parameter controlled how many misclassified examples the model will tolerate by adjusting the margin size. The model was evaluated using fMRI features only, sMRI features only, and the combination of fMRI and sMRI features.

Identifying easy cases

We randomly selected subsamples of the dataset (retaining 80% of participants in each subsample) to replicate the SVM training 100 times. For each 80% subsample, a separate SVM model was trained to predict the clinical labels (CN or AD), see Figure 6B. Note that the optimal C parameter was estimated once using the whole available

sample, as described above, and used across all subsamples. This was done to avoid creating major uncontrolled algorithmic variations. The linear discriminating weights of the SVM were still optimized independently for each subsample. Predictions of clinical labels were then made on the remaining 20% of subjects, that were not used for training. For a given individual, the hit probability was calculated as the frequency of correct clinical classification across all available SVM replications where the test set included that individual. Easy cases were defined as individuals with 100

Predicting easy cases

We trained a logistic regression classifier Fan et al. (2008a) to predict the AD easy cases. The logistic regression was trained using a L1 regularization on the coefficients, see Figure 6B. Class weight was balanced inversely proportional to class frequencies in the input data. A stratified shuffle split (100 splits, 20% test size) was used to estimate the performance of the model for the grid search of the hyper-parameter C (grid was $10^{-0.2}$ to 10^{1} with 15 equal steps). In this case, the *C* parameter controlled the sparseness of the weights.

Cross-validation

A nested cross-validation was performed for accuracy estimation and parameters optimization. The outer loop used to estimate the generalizability of the framework was a ten-fold cross-validation scheme. Each training fold included the full multi-stage process of subtype extraction, SVM prediction of clinical labels, identification of HPS and prediction of HPS with logistic regression. Sensitivity (true positive rate, TP), specificity (true negative rate, TN) and precision $(TP/(TP+(1-TN)))$ of the diagnosis were estimated across all test folds, in the AD vs CN prediction. Cross-validation nested inside the outer loop was used to search for the optimal hyper-parameters, Figure 6C.

Highly predictive signature

The HPS was obtained by considering all subtypes associated with non-zero weights by the sparse logistic regression model in Figure 6C stage 2. All nonzero weights were considered as part of the signature and we used the corresponding map associated with each subtype weight.

Prediction of progression to dementia

The easy cases model was used to identify MCI patients who have a HPS of AD dementia. The imaging sample for this experiment included the baseline structural and functional scans of all patients with MCI in the ADNI2 cohort, with at least 36 months of follow-up ($N = 56$). We further stratified the patients with MCI into stable MCI (sMCI, $N = 37$)), i.e. latest clinical status is MCI, and progressors (pMCI, $N = 19$), i.e. individuals whose most recent known clinical status is AD dementia, with progression from MCI to AD dementia occurring within 36 months. Note that no AV45 imaging data or genetic data, nor any data from the MCI cohort, were used to build the HPS model.

Statistical test of differences in model performance

We generated a confidence interval on the performance (i.e. precision, specificity and sensitivity) of a given model using a Monte-Carlo simulation. Taking the observed sensitivity and specificity, and using similar sample size to our experiment, we replicated the number of true and false positive detection 100000 times using independent Bernoulli variables, and derived replications of precision, specificity and sensitivity. By comparing these replications to the sensitivity, specificity and precision observed in other models, we estimated a p-value for differences in model performance (Phipson and Smyth 2010). A p-value smaller than 0.05 was interpreted as evidence of a significant difference in performance, and 0.001 as a strong evidence. This approach was first used in Figure 3 to contrast the performance of the HPS model to the baseline (SVM) model, both for AD vs CN and MCI progressor vs stable, as well as contrasting the performance of multimodal (fMRI+sMRI) model vs models using only fMRI or sMRI features. The same approach

was used to contrast our proposed model for MCI progressor vs stable with results from the literature, in Table 1. Note that, reflecting our hypotheses regarding the behaviour of the HPS model, the tests were one-sided for increase in specificity and precision, and one-sided for decrease in sensitivity.

Statistical test of enrichment

The HPS model was used to select a subset of the MCI population. We tested statistically if this subgroup was enriched for (1) progression to dementia; (2) AV45+, and; (3) ApoE4+. We implemented for this purpose a Monte-Carlo simulation, where we selected 100000 random subgroups out of the original MCI sample. By comparing the proportion of progressors (respectively AV45+ and ApoE4+) in these null replications to the actual observed values in the HPS subgroup, we estimated a p-value (Phipson and Smyth 2010) (one sided for increase). A p-value smaller than 0.05 was interpreted as evidence of a significant enrichment, and 0.001 as a strong evidence.

Public code and data

The code used in this experiment is available on a GitHub repository at the following URL[4]. An IPython Notebook is also provided with all of the figure generation scripts. Scikit-learn Pedregosa et al. (2011) version 0.18 was used for most of the machine learning algorithms and Nilearn Abraham et al. (2014) version 0.2.6 for visualization purposes.

ADNI dataset

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic

[4] `https://github.com/simexp/hpc`

resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD).

Supplementary Material – A brain signature highly predictive of future progression to Alzheimer's dementia

Table S2: Performance of the models. Prec: precision, Spec: specificity, Sens: sensitivity and N: number of selected subjects.

Modality	Algo	Contrast	Prec (%)	Spec (%)	Sens (%)	N
fMRI	Base	CN/AD	38.10	46.94	66.67	42
	HPS		60	95.92	12.5	5
sMRI	Base		66.67	83.67	66.67	24
	HPS		87.50	97.96	29.17	8
fMRI+sMRI	Base		69.57	85.71	66.67	23
	HPS		100	100	37.50	9
fMRI+sMRI	Base	sMCI/pMCI	73.33	89.19	57.89	15
	HPS		90	97.3	47.37	10

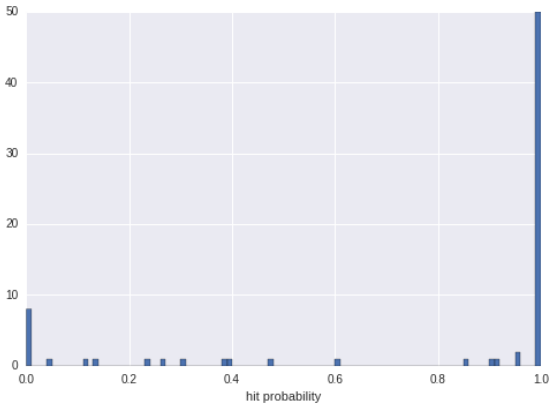

Figure S1: Hit-probability distribution obtained from replicating the SVM training 100 times from 80% of the training set.

CHAPTER 5

DISCUSSION

5.1 Contribution

5.1.1 Multisite

5.1.1.1 Multisite feasibility

In this thesis, I have shown that heterogeneity present in medical imaging negatively impacts our capacity to discriminate between various clinical statuses and that it is important to learn to deal with that heterogeneity to have actionable information. We first devised realistic Monte-Carlo simulations to model the impact of multisite data aggregation on statistical and prediction analysis. We have shown that the negative impact of multisite aggregation tends to diminish as we increase the sample size compared to a single site analysis. Those findings showed the viability of data aggregation and pushed us to investigate an even more ambitious goal: the generalizability of a model on sites never seen at training.

5.1.1.2 Multisite generalizability

Our second study explored the prediction generalizability of various multisite training schemes using real clinical effects. We demonstrated that training a model on a more diverse source of datasets leads to improved generalizability performance when tested on new sites not seen at training. In line with those findings, we have shown that training on one site and testing on the same site over estimates the model performance, in line with works comparing the subject and record levels (Little et al. 2017, Saeb et al. 2016).

Our findings regarding the use of multisite data to improve generalizability is concordant with a previous report that classified gender as a proof-of-concept application (Huf et al. 2014), and underscores the benefit of pooling multisite data for the purpose of generalizability and clinical use. Even though the accuracy scores reported in our

study are low compared to what has been shown previously in the literature, the findings are in line with what was found by others (Abraham et al. 2016, Varoquaux et al. 2017). We also have to keep in mind that the goal of our study was not to achieve a state of the art accuracy but rather to use a suitable dataset that can highlight the changes in performance from one training configuration with respect to others.

The benefits of multisite studies are twofold. First, the most obvious one, the sample size does increase at no extra cost when the multisite sample is the result of a data sharing effort. This increase in sample size however comes with an increase in the variance of the dataset. This variance is problematic since it decreases our ability to detect differences with small samples. However, it also comes with a second, less obvious benefit, as multisite heterogeneous samples better reflect the type of variability that one would encounter in an unseen dataset (a typical configuration in a clinical setup). Consequently accurately modeling the variability of the data (given enough samples through data aggregation) and teasing out the sources of variance that are not related to the pathology become highly beneficial for a machine learning model.

5.1.1.3 Importance of data aggregation

Combining the results from my first two articles, it is now obvious that data aggregation not only can be done but will also improve our model's ability to generalize to unseen data. I, therefore, would like to stress the importance of data aggregation as a major ingredient of future discoveries in medical imaging. It is only with a better view of the true heterogeneity that exists in neuroimaging, that we can hope to train generalizable and reliable models that will change the current "status quo" of medical diagnosis. This approach happened not long ago in the computer vision field with the advent of ImageNet (Deng et al. 2009), which completely changed the field and also popularized the use of deep learning. Initially, a lot of reticence came from the community stating that simpler problems could not be solved correctly with a few well-curated classes and therefore an extremely large dataset of a thousand classes, like ImageNet, would just render that task more difficult. What actually happened was that pre-trained models with ImageNet performed much better when fine-tuned on a specialized dataset than

models trained only on the specialized dataset (Everingham et al. 2010). What can be learned from this experience is that we do not need to have an extremely large dataset of one particular task or disease to learn a meaningful data reduction scheme that extracts relevant features. That being said, the quality of the data still needs to be assessed and controlled using quality assurance metrics. To devise similar experiments in the neuroimaging world, we could use multiple populational datasets with rich phenotypic data and try to predict those phenotypes or use multiple diseases combined together to form a large dataset. Rahim and colleagues Rahim et al. (2017) have shown that by learning to predict clinical labels from two different clinical populations, namely Schizophrenia and AD dementia, they were able to achieve better accuracy on an independent dataset of AD subjects. This proof of concept could be extended to a much larger database of diseases in order to expand on that idea.

5.1.1.4 Data sharing

It is now clearer as to what role data aggregation will play in the future of medicine and research. However, one problem remains regarding accessing and sharing the data in question. Some legal and ethical difficulties arise regarding that question and have unfortunately slowed down the data sharing process. Sharing should be performed with the best data anonymization practices. Nevertheless, there are recurrent concerns about the misuse of that medical data, for example by an insurance company that could subsequently have a detrimental impact on the population, e.g. refuse coverage or increase premiums based on the analysis of medical information of a particular subject. This is not directly related to data sharing, but rather more to scientific discovery and the use of those discoveries to find an individual's risk factors or prevalence to develop various diseases. Even though this issue has nothing to do with data sharing, some individuals use that argument to restrain it. My personal take on this is that it is a legislative problem and that it would not be as much of a concern given a legal prohibition to use any of that data to the detriment or to deny service/coverage to an individual. There is also the other issue of sharing licenses that preclude commercial use of a dataset or any derivative product. Currently, a lot of medical imaging datasets are released with such restrictive licenses.

This, I think, is a bit problematic from the perspective where we want to rapidly transfer and convert new academic findings into a tangible product, like clinical diagnoses and interventions. We will need to let private corporations use the data to train their model too or else we incur the risk of seeing those technologies trapped in academic labs.

5.1.2 Highly predictable cases

Regarding the heterogeneity of clinical labels, we have proposed a new method to automatically identify a subtype signature that can be found in a majority of AD dementia subjects and that is also found in a majority of MCI prior to their progression to AD dementia. By using this type of approach we have the ability to identify subtypes of patients that may refine our understanding of the clinical heterogeneity of known clinical labels (e.g. fronto-temporal dementia, AD dementia, Lewy body etc.). The outcome of this refinement is more precise and accurate models that will be trained to recognize those specific signatures of refined categories, which will lead to interventions tailored to those subgroups of individuals.

5.1.2.1 Multiple disease interaction

Another point that could partially explain some of the label heterogeneity is the comorbid effect of several major diseases, such as AD, cerebrovascular disease, Lewy body disease, or a frontotemporal degeneration that are usually responsible for dementia. These diseases are not exclusive and it is, therefore, possible to have more than one occurring at the same time as explained by Jellinger et al. (2014), Rabinovici et al. (2017). Older adults are also known to take multiple drugs to treat a variety of illnesses and symptoms resulting in potential interactions between these drugs and the brain. Some diseases currently do not have treatment although a lot of the comorbid factors, if untangled from other pathologies, may be treatable. The work that I have done related to label heterogeneity could be helpful to identify subgroups and find what other comorbid or protective factors (education level, brain reserve and cognitive reserve) are common to those subtypes, in the hope that it could draw a clearer map of the variability among

subjects and the potential treatment pathways.

Obviously, this kind of approach does not fit all applications, as it is most useful in cases where an action is deemed to be at high-risk or costly and an inaction not so much. In this type of context maximizing our chance of success is of the utmost importance. A potential use of this technique is in high-risk applications like finance (stock trading), for example, where you would like to buy and sell a stock only when it is very likely that the stock will rise or fall at the next time point and do nothing when you are unsure. Another application is in Ad Click Prediction (McMahan et al. 2013) where we need to decide which ad to show to which customer. If you identify a subgroup of subjects that are at high-risk of buying your product, you would be more inclined to bid more money to show your ad to those customers. Another example is loan default prediction: a bank will want to reject a loan automatically if the individual is at high-risk of default. Depending on what characterizes the non-HPC (e.g. noise or underrepresented subgroup) it may be possible to make the most of the HPC subjects, given a good denoising procedure or enough subjects and assuming that the measures are sensitive enough to the effect that we want to observe.

5.1.2.2 Classical confidence estimate

Finally, one could ask what is the difference between the hit-probability and the classic confidence estimate given by the model? Actually, the two could be used for similar objectives. One technical difference is that our hit-probability is calculated in a non-parametric way from a Monte-Carlo sampling procedure. The confidence estimate, on the other hand, is obtained from a parametric approximation. Our method has the advantage of being agnostic to the type of model evaluated (could be linear or non-linear) and the sampling strategy gives us a measure of the prediction robustness under data attrition. The confidence estimate gives the confidence level of the model into the prediction and the hit-probability is the likelihood of a subject to be correctly classified. The sampling strategy used to compute the hit-probability could also be investigated further. Currently, the sampling is a random split although we could imagine various sampling ratios or iteratively removing a group of subjects from the training set that

have been identified using a clustering strategy.

5.1.3 Other works and contributions

Some of my Ph.D. results have led to my contributions to other projects and those studies have helped me solidify ideas on some aspects of the heterogeneity challenge. In Tam et al. (2015) (see appendix I) we have identified several functional regions that could be used as biomarkers to discriminate between healthy controls and MCI in a multisite setup, and this highlighted the heterogeneity of the effect of MCI on brain connectivity across site and scanners. In Badhwar et al. (2017) (see appendix II) a meta-analysis on the functional regions reported to be affected by AD dementia in the literature showed a lot of variability between the reported regions, although some networks were consistently found like the default-mode network. Finally Orban et. al. in preparation (see appendix III) shows multiple functional subtypes that can be found independently across clinical AD datasets, and that these subtypes can be found in asymptomatic older adults at familial risk of AD.

5.2 Future works

5.2.1 Generative model for data augmentation

In line with our simulation work, we could explore the use of generative adversarial networks (GANs) to perform data augmentation as proposed by Goodfellow et al. (2014). Data augmentation is the process of creating supplemental examples from a set of original ones that have undergone some modifications like adding noise or lighting modification so that the model can learn to become invariant to those source of variance. The use of GANs or any other generative models would allow us to generate a realistic simulation of various pathology or artifacts. I have also advocated for data sharing and the use of multisite data initiatives, and I think that the use of GANs could be a solution to extract relevant information and share them while keeping a certain level of privacy. This approach was recently applied to blood tests Beaulieu-Jones et al. (2017). In order to train the discriminator under differential privacy constraint, Beaulieu-Jones

et al. (2017) added noise to the stochastic gradient descent process as outlined in Abadi et al. (2016). This type of approach has two advantages. First, it facilitates sharing of datasets due to the fact that the confidentiality of the training samples is preserved, since the generated samples no longer represent individual characteristics, while accurately approximating a group distribution. Second, it provides us with a generative model that we can use to perform data augmentation.

5.2.2 Direct application of HPC

One of the most interesting direct applications of the HPC work is in population enrichment for pharmaceutical clinical trials in Alzheimer's disease (Mathotaarachchi et al. 2017, Woo et al. 2017). By selecting subjects based on the presence of a specific signature it will be possible to increase the likelihood of exposing a drug to prodromal subjects at high-risk of progression to AD dementia. This strategy increases the effect size of the drug while reducing the sample size needed for a given effect and therefore reduces the cost a clinical trial. The signature shown in this study is specific enough to not be shared among all the AD dementia population (making it a subtype) but large enough to represent a substantial portion of the AD dementia population (more than half of AD dementia subjects and MCI progressors are positive for the signature). This method brings us closer to precision-medicine by proposing a middle ground between clinical group medicine and individual medicine.

5.2.2.1 Extension of the current HPC work

An extension of the HPC work would be to use more complex models (deep neural network and/or higher order models), in order to be able to model multiple modes of the distribution. By doing so we would be able to identify several HPC subtypes in one shot. We have to keep in mind that if the first stage uses a complex model, we need to use a similar model at the second stage, although the reverse is not true. Unfortunately, by using more complex models we would lose some interpretability of the models, which is a nice property of the current framework.

5.2.2.2 True supervisor

The HPC as it was proposed in my third paper is probably a good framework only for relatively small samples but will not scale well in its current format for large samples since we need to load all the data at once in memory to perform the analysis. It would, therefore, be useful to train that method in an online fashion with batches of data. The procedure could be reformulated as a two models problem. The first model would try to learn the task by estimating the targets and the second model (supervisor) would learn to predict what samples are likely to be correctly classified by the first model. That way both models could be trained simultaneously. This work is currently in progress and has shown promising results (on the MNIST dataset) that have not been included in my thesis.

5.3 Conclusion

The objective of this thesis was to explore the heterogeneity found in medical imaging datasets, and in particular in the context of data aggregation and pathological classification. My results have shown that data aggregation is not only possible but will also help in the generalization of the classification model to unseen data. It is therefore very important to leverage the heterogeneity found in the data to obtain more robust prediction models. I also have shown that label heterogeneity is a real problem that impairs our ability to make accurate predictions. I have proposed a solution to address this issue by identifying subgroups that can be reliably predicted with high accuracy. This type of approach could spur a very broad range of applications in the medical field where we have rich datasets with a limited number of examples and where ground truth labels are not well characterized. I also have discussed other potential applications of this approach and potential extensions of the proposed framework.

REFERENCES

Martín Abadi, Andy Chu, Ian Goodfellow, H Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang. Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, pages 308–318. ACM, 2016.

Alexandre Abraham, Fabian Pedregosa, Michael Eickenberg, Philippe Gervais, Andreas Mueller, Jean Kossaifi, Alexandre Gramfort, Bertrand Thirion, and Gael Varoquaux. Machine learning for neuroimaging with scikit-learn. *Frontiers in Neuroinformatics*, 8:14, 2014. ISSN 1662-5196. doi: 10.3389/fninf.2014.00014. URL `http://journal.frontiersin.org/article/10.3389/fninf.2014.00014`.

Alexandre Abraham, Michael Milham, Adriana Di Martino, R Cameron Craddock, Dimitris Samaras, Bertrand Thirion, and Gaël Varoquaux. Deriving reproducible biomarkers from multi-site resting-state data: An autism-based example. *NeuroImage*, 2016.

Y. Ad-Dab'bagh, D. Einarson, O. Lyttelton, J. S. Muehlboeck, K. Mok, O. Ivanov, R. D. Vincent, C. Lepage, J. Lerch, E. Fombonne, and A. C. Evans. The CIVET Image-Processing environment: A fully automated comprehensive pipeline for anatomical neuroimaging research. In M. Corbetta, editor, *Proceedings of the 12th Annual Meeting of the Human Brain Mapping Organization*, Florence, Italy, 2006a. Neuroimage.

Y. Ad-Dab'bagh, D. Einarson, O. Lyttelton, J. S. Muehlboeck, K. Mok, O. Ivanov, R. D. Vincent, C. Lepage, J. Lerch, E. Fombonne, and A. C. Evans. The CIVET Image-Processing Environment: A Fully Automated Comprehensive Pipeline for Anatomical Neuroimaging Research. In M. Corbetta, editor, *Proceedings of the 12th Annual Meeting of the Human Brain Mapping Organization*, Florence, Italy, 2006b. Neuroimage.

Stanisław Adaszewski, Juergen Dukart, Ferath Kherif, Richard Frackowiak, Bogdan Draganski, Alzheimer's Disease Neuroimaging Initiative, et al. How early can we

predict alzheimer's disease using computational anatomy? *Neurobiology of aging*, 34 (12):2815–2826, 2013.

A. Anand, Y. Li, Y. Wang, J. Wu, S. Gao, L. Bukhari, V. P. Mathews, A. Kalnin, and M. J. Lowe. Activity and connectivity of brain mood regulating circuit in depression: a functional magnetic resonance study. *Biological psychiatry*, 57(10):1079–1088, May 2005. ISSN 0006-3223. doi: 10.1016/j.biopsych.2005.02.021. URL `http://dx.doi.org/10.1016/j.biopsych.2005.02.021`.

Amit Anand, Yu Li, Yang Wang, Kathryn Gardner, and Mark J Lowe. Reciprocal effects of antidepressant treatment on activity and connectivity of the mood regulating circuit: an fmri study. *The Journal of neuropsychiatry and clinical neurosciences*, 2007.

Mohammad R Arbabshirani, Sergey Plis, Jing Sui, and Vince D Calhoun. Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls. *NeuroImage*, 145:137–165, 2017.

Rhoda Au, Ryan J Piers, and Lee Lancashire. Back to the future: Alzheimer's disease heterogeneity revisited. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, 1(3):368, 2015.

AmanPreet Badhwar, Angela Tam, Christian Dansereau, Pierre Orban, Felix Hoffstaedter, and Pierre Bellec. Resting-state network dysfunction in alzheimer's disease: A systematic review and meta-analysis. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, 8:73–85, 2017.

Thomas G Beach, Sarah E Monsell, Leslie E Phillips, and Walter Kukull. Accuracy of the clinical diagnosis of alzheimer disease at national institute on aging alzheimer disease centers, 2005–2010. *Journal of neuropathology & experimental neurology*, 71 (4):266–273, 2012.

Brett K Beaulieu-Jones, Zhiwei Steven Wu, Chris Williams, and Casey S Greene. Privacy-preserving generative deep neural networks support clinical data sharing. *bioRxiv*, page 159756, 2017.

P. Bellec, F. Carbonell, V. Perlbarg, and A. C. Evans. *A neuroimaging analysis kit for Octave and Matlab*, 2010a. URL http://code.google.com/p/niak/.

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 In Press+, 2011.

Pierre Bellec, Pedro Rosa-Neto, Oliver C. Lyttelton, Habib Benali, and Alan C. Evans. Multi-level bootstrap analysis of stable clusters in resting-state fMRI. *NeuroImage*, 51(3):1126–1139, July 2010b. ISSN 1095-9572. doi: 10.1016/j.neuroimage.2010.02.082. URL http://dx.doi.org/10.1016/j.neuroimage.2010.02.082.

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. *Frontiers in neuroinformatics*, 6, 2012. ISSN 1662-5196. doi: 10.3389/fninf.2012.00007. URL http://dx.doi.org/10.3389/fninf.2012.00007.

Pierre Bellec, Yassine Benhajali, Felix Carbonell, Christian Dansereau, Geneviève Albouy, Maxime Pelland, Cameron Craddock, Oliver Collignon, Julien Doyon, Emmanuel Stip, et al. Impact of the resolution of brain parcels on connectome-wide association studies in fmri. *NeuroImage*, 123:212–228, 2015.

Y. Benjamini and Y. Hochberg. Controlling the false-discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B*, 57:289–300, 1995.

B. Biswal, F. Z. Yetkin, V. M. Haughton, and J. S. Hyde. Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magnetic resonance in medicine*, 34(4):537–541, October 1995. ISSN 0740-3194. URL http://view.ncbi.nlm.nih.gov/pubmed/8524021.

Bharat B. Biswal, Maarten Mennes, Xi-Nian N. Zuo, Suril Gohel, Clare Kelly, Steve M. Smith, Christian F. Beckmann, Jonathan S. Adelstein, Randy L. Buckner, Stan Colcombe, Anne-Marie M. Dogonowski, Monique Ernst, Damien Fair, Michelle Hampson, Matthew J. Hoptman, James S. Hyde, Vesa J. Kiviniemi, Rolf Kötter, Shi-Jiang J. Li, Ching-Po P. 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. Rombouts, Bart Rypma, Bradley L. Schlaggar, Sein Schmidt, Rachael D. Seidler, Greg J. Siegle, Christian Sorg, Gao-Jun J. Teng, Juha Veijola, Arno Villringer, Martin Walter, Lihong Wang, Xu-Chu C. Weng, Susan Whitfield-Gabrieli, Peter Williamson, Christian Windischberger, Yu-Feng F. Zang, Hong-Ying Y. Zhang, F. Xavier Castellanos, and Michael P. Milham. Toward discovery science of human brain function. *Proceedings of the National Academy of Sciences of the United States of America*, 107(10):4734–4739, March 2010. ISSN 1091-6490. doi: 10.1073/pnas.0911855107. URL `http://dx.doi.org/10.1073/pnas.0911855107`.

Michael J Black and Anand Rangarajan. On the unification of line processes, outlier rejection, and robust statistics with applications in early vision. *International journal of computer vision*, 19(1):57–91, 1996.

Gregory G Brown, Daniel H Mathalon, Hal Stern, Judith Ford, Bryon Mueller, Douglas N Greve, Gregory McCarthy, James Voyvodic, Gary Glover, Michele Diaz, et al. Multisite reliability of cognitive bold data. *Neuroimage*, 54(3):2163–2175, 2011.

Rachel F Buckley, Aaron P Schultz, Trey Hedden, Kathryn V Papp, Bernard J Hanseeuw, Gad Marshall, Jorge Sepulcre, Emily E Smith, Dorene M Rentz, Keith A Johnson, et al. Functional network integrity presages cognitive decline in preclinical alzheimer disease. *Neurology*, pages 10–1212, 2017.

Carlos Cabral, Pedro M Morgado, Durval Campos Costa, Margarida Silveira, AlzheimerŒş s Disease Neuroimaging Initiative, et al. Predicting conversion from

mci to ad with fdg-pet brain images at different prodromal stages. *Computers in biology and medicine*, 58:101–109, 2015.

Vincent Camus, Pierre Payoux, Louisa Barré, Béatrice Desgranges, Thierry Voisin, Clovis Tauber, Renaud La Joie, Mathieu Tafani, Caroline Hommet, Gaël Chételat, et al. Using pet with 18f-av-45 (florbetapir) to quantify brain amyloid load in a clinical environment. *European journal of nuclear medicine and molecular imaging*, 39(4): 621–631, 2012.

Zografos Caramanos, Vladimir S Fonov, Simon J Francis, Sridar Narayanan, G Bruce Pike, D Louis Collins, and Douglas L Arnold. Gradient distortions in mri: Characterizing and correcting for their effects on siena-generated measures of brain volume change. *NeuroImage*, 49(2):1601–1611, 2010.

F. Carbonell, P. Bellec, and A. Shmuel. Quantification of the impact of a confounding variable on functional connectivity confirms anti-correlated networks in the resting-state. *Neuroimage*, 86:343–353, Feb 2014. doi: 10.1016/j.neuroimage.2013.10.013. URL `http://dx.doi.org/10.1016/j.neuroimage.2013.10.013`.

Mustafa S Çetin, Fletcher Christensen, Christopher C Abbott, Julia M Stephen, Andrew R Mayer, José M Cañive, Juan R Bustillo, Godfrey D Pearlson, and Vince D Calhoun. Thalamus and posterior temporal lobe show greater inter-network connectivity at rest and across sensory paradigms in schizophrenia. *Neuroimage*, 97:117–126, 2014.

Edward Challis, Peter Hurley, Laura Serra, Marco Bozzali, Seb Oliver, and Mara Cercignani. Gaussian process classification of alzheimer's disease and mild cognitive impairment from resting-state fmri. *NeuroImage*, 112:232–243, 2015.

Chih-Chung Chang and Chih-Jen Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011. Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

Gang Chen, B. Douglas Ward, Chunming Xie, Wenjun Li, Zhilin Wu, Jennifer L. Jones, Malgorzata Franczak, Piero Antuono, and Shi-Jiang Li. Classification of Alzheimer Disease, Mild Cognitive Impairment, and Normal Cognitive Status with Large-Scale Network Analysis Based on Resting-State Functional MR Imaging. *Radiology*, 259 (1):213–221, April 2011. doi: 10.1148/radiol.10100734. URL `http://dx.doi.org/10.1148/radiol.10100734`.

Gang Chen, Ziad S Saad, Jennifer C Britton, Daniel S Pine, and Robert W Cox. Linear mixed-effects modeling approach to fmri group analysis. *Neuroimage*, 73:176–190, 2013.

Bo Cheng, Mingxia Liu, Heung-Il Suk, Dinggang Shen, Daoqiang Zhang, AlzheimerâĂŹs Disease Neuroimaging Initiative, et al. Multimodal manifold-regularized transfer learning for mci conversion prediction. *Brain imaging and behavior*, 9(4):913–926, 2015a.

Bo Cheng, Mingxia Liu, Daoqiang Zhang, Brent C Munsell, and Dinggang Shen. Domain transfer learning for mci conversion prediction. *IEEE Transactions on Biomedical Engineering*, 62(7):1805–1817, 2015b.

Wei Cheng, Lena Palaniyappan, Mingli Li, Keith M Kendrick, Jie Zhang, Qiang Luo, Zening Liu, Rongjun Yu, Wei Deng, Qiang Wang, Xiaohong Ma, Wanjun Guo, Susan Francis, Peter Liddle, Andrew R Mayer, Gunter Schumann, Tao Li, and Jianfeng Feng. Voxel-based, brain-wide association study of aberrant functional connectivity in schizophrenia implicates thalamocortical circuitry. *Npj Schizophrenia*, 1:–, May 2015c. URL `http://dx.doi.org/10.1038/npjschz.2015.16`.

G. Chetelat, B. Desgranges, V. de la Sayette, F. Viader, F. Eustache, and J-C. Baron. Mild cognitive impairment: Can fdg-pet predict who is to rapidly convert to alzheimer's disease? *Neurology*, 60(8):1374–1377, Apr 2003.

Z. H. Cho, S. C. Chung, D. W. Lim, and E. K. Wong. Effects of the acoustic noise of

the gradient systems on fmri: a study on auditory, motor, and visual cortices. *Magn Reson Med*, 39(2):331–335, Feb 1998.

Jacob Cohen. A power primer. *Psychological bulletin*, 112(1):155, 1992.

D. L. Collins, P. Neelin, T. M. Peters, and A. C. Evans. Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. *Journal of computer assisted tomography*, 18(2):192–205, 1994. ISSN 0363-8715. URL http://view.ncbi.nlm.nih.gov/pubmed/8126267.

Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine Learning*, 20 (3):273–297, September 1995. ISSN 0885-6125. doi: 10.1007/BF00994018. URL http://dx.doi.org/10.1007/BF00994018.

Sergi G Costafreda, Akash Khanna, Janaina Mourao-Miranda, and Cynthia H Y Fu. Neural correlates of sad faces predict clinical remission to cognitive behavioural therapy in depression. *Neuroreport*, 20(7):637–641, May 2009. doi: 10.1097/WNR.0b013e3283294159. URL http://dx.doi.org/10.1097/WNR.0b013e3283294159.

Kathryn R Cullen, Dylan G Gee, Bonnie Klimes-Dougan, Vilma Gabbay, Leslie Hulvershorn, Bryon A Mueller, Jazmin Camchong, Christopher J Bell, Alaa Houri, Sanjiv Kumra, Kelvin O Lim, F. Xavier Castellanos, and Michael P Milham. A preliminary study of functional connectivity in comorbid adolescent depression. *Neurosci Lett*, 460(3):227–231, Sep 2009. doi: 10.1016/j.neulet.2009.05.022. URL http://dx.doi.org/10.1016/j.neulet.2009.05.022.

Marcello D'Amelio and Paolo Maria Rossini. Brain excitability and connectivity of neuronal assemblies in alzheimer's disease: from animal models to human findings. *Progress in neurobiology*, 99(1):42–60, October 2012. ISSN 1873-5118. URL http://view.ncbi.nlm.nih.gov/pubmed/22789698.

J. S. Damoiseaux, S. A. R. B. Rombouts, F. Barkhof, P. Scheltens, C. J. Stam, S. M. Smith, and C. F. Beckmann. Consistent resting-state networks across healthy subjects.

Proceedings of the National Academy of Sciences, 103(37):13848–13853, September 2006. ISSN 1091-6490. doi: 10.1073/pnas.0601417103. URL http://dx.doi.org/10.1073/pnas.0601417103.

Christian Dansereau, Celine Risterucci, Emilio Merlo Pich, Douglas Arnold, and Pierre Bellec. A power analysis for multisite studies in resting-state functional connectivity, with an application to clinical trials in alzheimer's disease. volume 9, pages P248 – P249, 2013. doi: http://dx.doi.org/10.1016/j.jalz.2013.05.489. URL http://www.sciencedirect.com/science/article/pii/S1552526013011461. Alzheimer's Association International Conference 2013 Alzheimer's Association International Conference 2013.

Christian Dansereau, Pierre Bellec, Kangjoo Lee, Francesca Pittau, Jean Gotman, and Christophe Grova. Detection of abnormal resting-state networks in individual patients suffering from focal epilepsy: An initial step toward individual connectivity assessment. *Frontiers in Neuroscience*, 8(419), 2014. ISSN 1662-453X. doi: 10.3389/fnins.2014.00419. URL http://www.frontiersin.org/brain_imaging_methods/10.3389/fnins.2014.00419/abstract.

Christian Dansereau, Yassine Benhajali, Celine Risterucci, Emilio Merlo Pich, Pierre Orban, Douglas Arnold, and Pierre Bellec. Statistical power and prediction accuracy in multisite resting-state fmri connectivity. *NeuroImage*, 149:220–232, 2017.

J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR09*, 2009.

John Desmond and Gary Glover. Estimating sample size in functional mri (fmri) neuroimaging studies: Statistical power analyses. *Journal of Neuroscience Methods*, 118 (2):115–128, August 2002. ISSN 01650270. URL http://dx.doi.org/10.1016/s0165-0270(02)00121-8.

Adriana Di Martino, Chao-Gan Yan, Qingyang Li, Erin Denio, Francisco X Castellanos, Kaat Alaerts, Jeffrey S Anderson, Michal Assaf, Susan Y Bookheimer, Mirella

Dapretto, et al. The autism brain imaging data exchange: towards large-scale evaluation of the intrinsic brain architecture in autism. *Molecular psychiatry*, 19(6):659, 2014.

Taciana G Costa Dias, Swathi P Iyer, Samuel D Carpenter, Robert P Cary, Vanessa B Wilson, Suzanne H Mitchell, Joel T Nigg, and Damien A Fair. Characterizing heterogeneity in children with and without adhd based on reward system connectivity. *Developmental cognitive neuroscience*, 11:155–174, 2015.

Aoyan Dong, Jon B Toledo, Nicolas Honnorat, Jimit Doshi, Erdem Varol, Aristeidis Sotiras, David Wolk, John Q Trojanowski, Christos Davatzikos, and AlzheimerâĂŹs Disease Neuroimaging Initiative. Heterogeneity of neuroanatomical patterns in prodromal alzheimerâĂŹs disease: links to cognition, progression and biomarkers. *Brain*, 140(3):735–747, 2016.

Andrew T Drysdale, Logan Grosenick, Jonathan Downar, Katharine Dunlop, Farrokh Mansouri, Yue Meng, Robert N Fetcho, Benjamin Zebley, Desmond J Oathes, Amit Etkin, et al. Resting-state connectivity biomarkers define neurophysiological subtypes of depression. *Nature medicine*, 23(1):28–38, 2017.

Bruno Dubois, Howard H. Feldman, Claudia Jacova, Steven T. Dekosky, Pascale Barberger-Gateau, Jeffrey Cummings, André Delacourte, Douglas Galasko, Serge Gauthier, Gregory Jicha, Kenichi Meguro, John O'brien, Florence Pasquier, Philippe Robert, Martin Rossor, Steven Salloway, Yaakov Stern, Pieter J. Visser, and Philip Scheltens. Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. *Lancet neurology*, 6(8):734–746, August 2007. ISSN 1474-4422. doi: 10.1016/S1474-4422(07)70178-3. URL `http://dx.doi.org/10.1016/S1474-4422(07)70178-3`.

Julien Dubois and Ralph Adolphs. Building a science of individual differences from fmri. *Trends in cognitive sciences*, 20(6):425–443, 2016.

Joke Durnez, Beatrijs Moerkerke, and Thomas E Nichols. Post-hoc power estimation for topological inference in fmri. *Neuroimage*, 84:45–64, 2014.

V. Edward, C. Windischberger, R. Cunnington, M. Erdler, R. Lanzenberger, D. Mayer, W. Endl, and R. Beisteiner. Quantification of fmri artifact reduction by a novel plaster cast head holder. *Hum Brain Mapp*, 11(3):207–213, Nov 2000.

M. R. Elliott, R. W. Bowtell, and P. G. Morris. The effect of scanner sound in visual, motor, and auditory functional mri. *Magn Reson Med*, 41(6):1230–1235, Jun 1999.

Simon F. Eskildsen, Pierrick Coupé, Daniel Garc'ıa-Lorenzo, Vladimir Fonov, Jens C. Pruessner, and D. Louis Collins. Prediction of Alzheimer's disease in subjects with mild cognitive impairment from the ADNI cohort using patterns of cortical thinning. *NeuroImage*, 65:511–521, January 2013. doi: 10.1016/j.neuroimage.2012.09.058. URL `http://dx.doi.org/10.1016/j.neuroimage.2012.09.058`.

M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The pascal visual object classes (voc) challenge. *International Journal of Computer Vision*, 88 (2):303–338, June 2010.

Damien A Fair, Joel T Nigg, Swathi Iyer, Deepti Bathula, Kathryn L Mills, Nico U F Dosenbach, Bradley L Schlaggar, Maarten Mennes, David Gutman, Saroja Bangaru, Jan K Buitelaar, Daniel P Dickstein, Adriana Di Martino, David N Kennedy, Clare Kelly, Beatriz Luna, Julie B Schweitzer, Katerina Velanova, Yu-Feng Wang, Stewart Mostofsky, F. Xavier Castellanos, and Michael P Milham. Distinct neural signatures detected for adhd subtypes after controlling for micro-movements in resting state functional connectivity mri data. *Front Syst Neurosci*, 6:80, 2012. doi: 10.3389/fnsys.2012.00080. URL `http://dx.doi.org/10.3389/fnsys.2012.00080`.

Rong-En Fan, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, and Chih-Jen Lin. Liblinear: A library for large linear classification. *The Journal of Machine Learning Research*, 9:1871–1874, 2008a.

Yong Fan, Dinggang Shen, and Christos Davatzikos. Classification of structural images via high-dimensional image warping, robust feature extraction, and svm. *Med Image Comput Comput Assist Interv*, 8(Pt 1):1–8, 2005.

Yong Fan, Nematollah Batmanghelich, Chris M Clark, Christos Davatzikos, AlzheimerâĂŹs Disease Neuroimaging Initiative, et al. Spatial patterns of brain atrophy in mci patients, identified via high-dimensional pattern classification, predict subsequent cognitive decline. *Neuroimage*, 39(4):1731–1743, 2008b.

Fang Fang, Scott O Murray, and Sheng He. Duration-dependent fmri adaptation and distributed viewer-centered face representation in human visual cortex. *Cereb Cortex*, 17(6):1402–1411, Jun 2007. doi: 10.1093/cercor/bhl053. URL `http://dx.doi.org/10.1093/cercor/bhl053`.

Daniel Feaster, Susan Mikulich-Gilbertson, and Ahnalee Brincks. Modeling site effects in the design and analysis of multi-site trials. *The American journal of drug and alcohol abuse*, 37(5):383–391, September 2011. ISSN 1097-9891. URL `http://dx.doi.org/10.3109/00952990.2011.600386`.

David A Feinberg, Steen Moeller, Stephen M Smith, Edward Auerbach, Sudhir Ramanna, Matthias Gunther, Matt F Glasser, Karla L Miller, Kamil Ugurbil, and Essa Yacoub. Multiplexed echo planar imaging for sub-second whole brain fmri and fast diffusion imaging. *PLoS One*, 5(12):e15710, 2010. doi: 10.1371/journal.pone.0015710. URL `http://dx.doi.org/10.1371/journal.pone.0015710`.

Michelle T Fodero-Tavoletti, Nobuyuki Okamura, Shozo Furumoto, Rachel S Mulligan, Andrea R Connor, Catriona A McLean, Diana Cao, Angela Rigopoulos, Glenn A Cartwright, Graeme OâĂŹKeefe, et al. 18f-thk523: a novel in vivo tau imaging ligand for alzheimerâĂŹs disease. *Brain*, 134(4):1089–1100, 2011.

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 2011.

ISSN 1095-9572. doi: 10.1016/j.neuroimage.2010.07.033. URL http://dx.doi.org/10.1016/j.neuroimage.2010.07.033.

Michael D. Fox and Marcus E. Raichle. Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nat Rev Neurosci*, 8 (9):700–711, September 2007. ISSN 1471-003X. doi: 10.1038/nrn2201. URL http://dx.doi.org/10.1038/nrn2201.

Michael D Fox, Dongyang Zhang, Abraham Z Snyder, and Marcus E Raichle. The global signal and observed anticorrelated resting state brain networks. *J Neurophysiol*, 101 (6):3270–3283, Jun 2009. doi: 10.1152/jn.90777.2008. URL http://dx.doi.org/10.1152/jn.90777.2008.

Jerome Friedman, Trevor Hastie, and Robert Tibshirani. *The elements of statistical learning*, volume 1. Springer series in statistics New York, 2001.

Lee Friedman and Gary Glover. Report on a multicenter fmri quality assurance protocol. *Journal of magnetic resonance imaging : JMRI*, 23(6):827–839, June 2006. ISSN 1053-1807. URL http://dx.doi.org/10.1002/jmri.20583.

Lee Friedman, Gary Glover, and The FBIRN Consortium. Reducing interscanner variability of activation in a multicenter fmri study: Controlling for signal-to-fluctuation-noise-ratio (sfnr) differences. *NeuroImage*, 33(2):471–481, November 2006. ISSN 10538119. URL http://dx.doi.org/10.1016/j.neuroimage.2006.07.012.

Lee Friedman, Hal Stern, Gregory G Brown, Daniel H Mathalon, Jessica Turner, Gary H Glover, Randy L Gollub, John Lauriello, Kelvin O Lim, Tyrone Cannon, et al. Test–retest and between-site reliability in a multicenter fmri study. *Human brain mapping*, 29(8):958–972, 2008.

Giovanni B. Frisoni, Alberto Redolfi, David Manset, Marc-Etienne Rousseau, Arthur Toga, and Alan C. Evans. Virtual imaging laboratories for marker discovery in neurodegenerative diseases. *Nature Reviews Neurology*, 7(8):429–438, July 2011. ISSN

1759-4758. doi: 10.1038/nrneurol.2011.99. URL http://dx.doi.org/10.1038/nrneurol.2011.99.

Cynthia H Y Fu, Janaina Mourao-Miranda, Sergi G Costafreda, Akash Khanna, Andre F Marquand, Steve C R Williams, and Michael J Brammer. Pattern classification of sad facial processing: toward the development of neurobiological markers in depression. *Biol Psychiatry*, 63(7):656–662, Apr 2008. doi: 10.1016/j.biopsych.2007.08.020. URL http://dx.doi.org/10.1016/j.biopsych.2007.08.020.

Kathleen M Gates, Peter CM Molenaar, Swathi P Iyer, Joel T Nigg, and Damien A Fair. Organizing heterogeneous samples using community detection of gimme-derived resting state functional networks. *PloS one*, 9(3):e91322, 2014.

Federico Giove, Tommaso Gili, Vittorio Iacovella, Emiliano Macaluso, and Bruno Maraviglia. Images-based suppression of unwanted global signals in resting-state functional connectivity studies. *Magnetic resonance imaging*, 27(8):1058–1064, October 2009. ISSN 1873-5894. doi: 10.1016/j.mri.2009.06.004. URL http://dx.doi.org/10.1016/j.mri.2009.06.004.

Gary H Glover, Bryon A Mueller, Jessica A Turner, Theo GM van Erp, Thomas T Liu, Douglas N Greve, James T Voyvodic, Jerod Rasmussen, Gregory G Brown, David B Keator, et al. Function biomedical informatics research network recommendations for prospective multicenter functional mri studies. *Journal of Magnetic Resonance Imaging*, 36(1):39–54, 2012.

Randy L Gollub, Jody M Shoemaker, Margaret D King, Tonya White, Stefan Ehrlich, Scott R Sponheim, Vincent P Clark, Jessica A Turner, Bryon A Mueller, Vince Magnotta, et al. The mcic collection: a shared repository of multi-modal, multi-site brain image data from a clinical investigation of schizophrenia. *Neuroinformatics*, 11(3):367–388, 2013.

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sher-

jil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

Michael D. Greicius, Gaurav Srivastava, Allan L. Reiss, and Vinod Menon. Default-mode network activity distinguishes Alzheimer's disease from healthy aging: Evidence from functional MRI. *Proceedings of the National Academy of Sciences of the United States of America*, 101(13):4637–4642, March 2004. ISSN 1091-6490. doi: 10.1073/pnas.0308627101. URL http://dx.doi.org/10.1073/pnas.0308627101.

Michael D. Greicius, Benjamin H. Flores, Vinod Menon, Gary H. Glover, Hugh B. Solvason, Heather Kenna, Allan L. Reiss, and Alan F. Schatzberg. Resting-state functional connectivity in major depression: abnormally increased contributions from subgenual cingulate cortex and thalamus. *Biological psychiatry*, 62(5):429–437, September 2007. ISSN 0006-3223. doi: 10.1016/j.biopsych.2006.09.020. URL http://dx.doi.org/10.1016/j.biopsych.2006.09.020.

Tim Hahn, Andre F Marquand, Ann-Christine Ehlis, Thomas Dresler, Sarah Kittel-Schneider, Tomasz A Jarczok, Klaus-Peter Lesch, Peter M Jakob, Janaina Mourao-Miranda, Michael J Brammer, and Andreas J Fallgatter. Integrating neurobiological markers of depression. *Arch Gen Psychiatry*, 68(4):361–368, Apr 2011. doi: 10.1001/archgenpsychiatry.2010.178. URL http://dx.doi.org/10.1001/archgenpsychiatry.2010.178.

John Hardy and Dennis J Selkoe. The amyloid hypothesis of alzheimer's disease: progress and problems on the road to therapeutics. *Science*, 297(5580):353–356, Jul 2002. doi: 10.1126/science.1072994. URL http://dx.doi.org/10.1126/science.1072994.

Egbert Hartstra, Simone KÃijhn, Tom Verguts, and Marcel Brass. The implementation of verbal instructions: an fmri study. *Hum Brain Mapp*, 32(11):1811–1824, Nov 2011. doi: 10.1002/hbm.21152. URL http://dx.doi.org/10.1002/hbm.21152.

David J Heeger and David Ress. What does fmri tell us about neuronal activity? *Nat Rev Neurosci*, 3(2):142–151, Feb 2002. doi: 10.1038/nrn730. URL `http://dx.doi.org/10.1038/nrn730`.

R. D. Hoge, J. Atkinson, B. Gill, G. R. Crelier, S. Marrett, and G. B. Pike. Investigation of bold signal dependence on cerebral blood flow and oxygen consumption: the deoxyhemoglobin dilution model. *Magn Reson Med*, 42(5):849–863, Nov 1999.

Wolfgang Huf, Klaudius Kalcher, Roland N Boubela, Georg Rath, Andreas Vecsei, Peter Filzmoser, and Ewald Moser. On the generalizability of resting-state fmri machine learning classifiers. *Frontiers in human neuroscience*, 8, 2014.

J. D. Hunter. Matplotlib: A 2d graphics environment. *Computing In Science & Engineering*, 9(3):90–95, 2007.

Quentin JM Huys, Tiago V Maia, and Michael J Frank. Computational psychiatry as a bridge from neuroscience to clinical applications. *Nature neuroscience*, 19(3):404, 2016.

Jihye Hwang, Chan Mi Kim, Seun Jeon, Jong Min Lee, Yun Jeong Hong, Jee Hoon Roh, and Jae-Hong Lee. Prediction of alzheimerâĂŹs disease pathophysiology based on cortical thickness patterns. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, 11(7):P541, 2015.

Kiho Im, Jong-Min Lee, Oliver Lyttelton, Sun Hyung Kim, Alan C Evans, and Sun I Kim. Brain size and cortical structure in the adult human brain. *Cerebral Cortex*, 18 (9):2181–2191, 2008.

Thomas R Insel. Rethinking schizophrenia. *Nature*, 468(7321):187, 2010.

C. R. Jack, R. C. Petersen, Y. C. Xu, P. C. O'Brien, G. E. Smith, R. J. Ivnik, B. F. Boeve, S. C. Waring, E. G. Tangalos, and E. Kokmen. Prediction of ad with mri-based hippocampal volume in mild cognitive impairment. *Neurology*, 52(7):1397–1403, Apr 1999.

Clifford R Jack, David S Knopman, William J Jagust, Leslie M Shaw, Paul S Aisen, Michael W Weiner, Ronald C Petersen, and John Q Trojanowski. Hypothetical model of dynamic biomarkers of the alzheimer's pathological cascade. *Lancet Neurol*, 9 (1):119–128, Jan 2010. doi: 10.1016/S1474-4422(09)70299-6. URL http://dx.doi.org/10.1016/S1474-4422(09)70299-6.

Kurt A Jellinger et al. Neuropathology of dementia disorders. *J. Alz. Dis. Parkinsonism*, 4:135, 2014.

Jaeseung Jeong. Eeg dynamics in patients with alzheimer's disease. *Clin Neurophysiol*, 115(7):1490–1505, Jul 2004. doi: 10.1016/j.clinph.2004.01.001. URL http://dx.doi.org/10.1016/j.clinph.2004.01.001.

Biao Jie, Daoqiang Zhang, Chong-Yaw Wee, and Dinggang Shen. Topological graph kernel on multiple thresholded functional connectivity networks for mild cognitive impairment classification. *Hum. Brain Mapp.*, 35(7):2876–2897, July 2014. ISSN 1097-0193. URL http://dx.doi.org/10.1002/hbm.22353.

Eric Jones, Travis Oliphant, Pearu Peterson, et al. SciPy: Open source scientific tools for Python, 2001–. URL http://www.scipy.org/. [Online; accessed <today>].

Jorge Jovicich, Ludovico Minati, Moira Marizzoni, Rocco Marchitelli, Roser Sala-Llonch, David Bartrés-Faz, Jennifer Arnold, Jens Benninghoff, Ute Fiedler, Luca Roccatagliata, et al. Longitudinal reproducibility of default-mode network connectivity in healthy elderly participants: A multicentric resting-state fmri study. *NeuroImage*, 124:442–454, 2016.

Shitij Kapur, Anthony G Phillips, and Thomas R Insel. Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? *Molecular psychiatry*, 17(12):1174, 2012.

GB Karas, P Scheltens, SARB Rombouts, PJ Visser, RA Van Schijndel, NC Fox, and F Barkhof. Global and local gray matter loss in mild cognitive impairment and alzheimer's disease. *Neuroimage*, 23(2):708–716, 2004.

Tobias Kaufmann, Dag Alnæs, Christine Lycke Brandt, Nhat Trung Doan, Karolina Kauppi, Francesco Bettella, Trine V Lagerberg, Akiah O Berg, Srdjan Djurovic, Ingrid Agartz, et al. Task modulations and clinical manifestations in the brain functional connectome in 1615 fmri datasets. *NeuroImage*, 147:243–252, 2017.

Yasuhiro Kawasaki, Michio Suzuki, Ferath Kherif, Tsutomu Takahashi, Shi-Yu Zhou, Kazue Nakamura, Mie Matsui, Tomiki Sumiyoshi, Hikaru Seto, and Masayoshi Kurachi. Multivariate voxel-based morphometry successfully differentiates schizophrenia patients from healthy controls. *Neuroimage*, 34(1):235–242, Jan 2007. doi: 10.1016/j.neuroimage.2006.08.018. URL `http://dx.doi.org/10.1016/j.neuroimage.2006.08.018`.

Ali Khazaee, Ata Ebrahimzadeh, and Abbas Babajani-Feremi. Identifying patients with alzheimerâĂŹs disease using resting-state fmri and graph theory. *Clinical Neurophysiology*, 126(11):2132–2141, 2015.

LA Kilpatrick, DH Zald, JV Pardo, and LF Cahill. Sex-related differences in amygdala functional connectivity during resting conditions. *Neuroimage*, 30(2):452–461, 2006.

June Sic Kim, Vivek Singh, Jun Ki Lee, Jason Lerch, Yasser Ad-Dab'bagh, David MacDonald, Jong Min Lee, Sun I Kim, and Alan C Evans. Automated 3-d extraction and evaluation of the inner and outer cortical surfaces using a laplacian map and partial volume effect classification. *Neuroimage*, 27(1):210–221, 2005.

Markus Klarhofer, Markus Barth, and Ewald Moser. Comparison of multi-echo spiral and echo planar imaging in functional mri. *Magn Reson Imaging*, 20(4):359–364, May 2002.

Alex Kogan, Kathryn Alpert, Jose Luis Ambite, Daniel S Marcus, and Lei Wang. Northwestern university schizophrenia data sharing for schizconnect: A longitudinal dataset for large-scale integration. *NeuroImage*, 124:1196–1201, 2016.

Igor O Korolev, Laura L Symonds, Andrea C Bozoki, Alzheimer's Disease Neuroimaging Initiative, et al. Predicting progression from mild cognitive impairment

to alzheimer's dementia using clinical, mri, and plasma biomarkers via probabilistic pattern classification. *PloS one*, 11(2):e0138866, 2016.

S. B. Kotsiantis. Supervised machine learning: A review of classification techniques. pages 3–24, 2007. URL http://dl.acm.org/citation.cfm?id=1566770.1566773.

Benjamin Lam, Mario Masellis, Morris Freedman, Donald T Stuss, and Sandra E Black. Clinical, imaging, and pathological heterogeneity of the alzheimer's disease syndrome. *Alzheimer's research & therapy*, 5(1):1, 2013.

Zhiqiang Lao, Dinggang Shen, Zhong Xue, Bilge Karacali, Susan M Resnick, and Christos Davatzikos. Morphological classification of brains via high-dimensional shape transformations and machine learning methods. *Neuroimage*, 21(1):46–57, Jan 2004.

Jason P. Lerch, Jens C. Pruessner, Alex Zijdenbos, Harald Hampel, Stefan J. Teipel, and Alan C. Evans. Focal Decline of Cortical Thickness in Alzheimer's Disease Identified by Computational Neuroanatomy. *Cerebral Cortex*, 15(7):995–1001, July 2005. ISSN 1047-3211. doi: 10.1093/cercor/bhh200. URL http://dx.doi.org/10.1093/cercor/bhh200.

Meng Liang, Yuan Zhou, Tianzi Jiang, Zhening Liu, Lixia Tian, Haihong Liu, and Yihui Hao. Widespread functional disconnectivity in schizophrenia with resting-state functional magnetic resonance imaging. *Neuroreport*, 17(2):209–213, Feb 2006.

Fa-Hsuan Lin, Teng-Yi Huang, Nan-Kuei Chen, Fu-Nien Wang, Steven M Stufflebeam, John W Belliveau, Lawrence L Wald, and Kenneth K Kwong. Functional mri using regularized parallel imaging acquisition. *Magn Reson Med*, 54(2):343–353, Aug 2005. doi: 10.1002/mrm.20555. URL http://dx.doi.org/10.1002/mrm.20555.

Max A Little, Gael Varoquaux, Sohrab Saeb, Luca Lonini, Arun Jayaraman, David C Mohr, and Konrad P Kording. Using and understanding cross-validation strategies. perspectives on saeb et al. *GigaScience*, 6(5):1–6, 2017.

Mingxia Liu, Daoqiang Zhang, and Dinggang Shen. View-centralized multi-atlas classification for alzheimer's disease diagnosis. *Human brain mapping*, 36(5):1847–1865, 2015.

Xin Liu, Duygu Tosun, Michael W Weiner, Norbert Schuff, Alzheimer's Disease Neuroimaging Initiative, et al. Locally linear embedding (lle) for mri based alzheimer's disease classification. *NeuroImage*, 83:148–157, 2013.

Yong Liu, Chunshui Yu, Meng Liang, Jun Li, Lixia Tian, Yuan Zhou, Wen Qin, Kuncheng Li, and Tianzi Jiang. Whole brain functional connectivity in the early blind. *Brain*, 130(8):2085–2096, August 2007. ISSN 1460-2156. doi: 10.1093/brain/awm121. URL http://dx.doi.org/10.1093/brain/awm121.

Mark J Lowe, Micheal D Phillips, Joseph T Lurito, David Mattson, Mario Dzemidzic, and Vincent P Mathews. Multiple sclerosis: low-frequency temporal blood oxygen level-dependent fluctuations indicate reduced functional connectivity initial results. *Radiology*, 224(1):184–192, Jul 2002. doi: 10.1148/radiol.2241011005. URL http://dx.doi.org/10.1148/radiol.2241011005.

Torben E. Lund, Kristoffer H. Madsen, Karam Sidaros, Wen-Lin Luo, and Thomas E. Nichols. Non-white noise in fMRI: does modelling have an impact? *NeuroImage*, 29 (1):54–66, January 2006. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2005.07.005. URL http://dx.doi.org/10.1016/j.neuroimage.2005.07.005.

Oliver Lyttelton, Maxime Boucher, Steven Robbins, and Alan Evans. An unbiased iterative group registration template for cortical surface analysis. *Neuroimage*, 34(4): 1535–1544, 2007.

David MacDonald, Noor Kabani, David Avis, and Alan C Evans. Automated 3-d extraction of inner and outer surfaces of cerebral cortex from mri. *NeuroImage*, 12(3): 340–356, 2000.

Niklas Marklund, Kaj Blennow, Henrik Zetterberg, Elisabeth Ronne-Engström, Per Enblad, and Lars Hillered. Monitoring of brain interstitial total tau and beta amyloid

proteins by microdialysis in patients with traumatic brain injury. *J. Neurosurg.*, 110 (6):1227–1237, June 2009.

Andre F Marquand, Janaina MourÃčo-Miranda, Michael J Brammer, Anthony J Cleare, and Cynthia H Y Fu. Neuroanatomy of verbal working memory as a diagnostic biomarker for depression. *Neuroreport*, 19(15):1507–1511, Oct 2008. doi: 10.1097/WNR.0b013e328310425e. URL http://dx.doi.org/10.1097/WNR.0b013e328310425e.

Sulantha Mathotaarachchi, Tharick A Pascoal, Monica Shin, Andrea L Benedet, Min Su Kang, Thomas Beaudry, Vladimir S Fonov, Serge Gauthier, Pedro Rosa-Neto, AlzheimerâĂŹs Disease Neuroimaging Initiative, et al. Identifying incipient dementia individuals using machine learning and amyloid imaging. *Neurobiology of Aging*, 2017.

Niklas Mattsson, Henrik Zetterberg, Oskar Hansson, Niels Andreasen, Lucilla Parnetti, Michael Jonsson, Sanna-Kaisa Herukka, Wiesje M van der Flier, Marinus A Blankenstein, Michael Ewers, Kenneth Rich, Elmar Kaiser, Marcel Verbeek, Magda Tsolaki, Ezra Mulugeta, Erik RosÃl'n, Dag Aarsland, Pieter Jelle Visser, Johannes SchrÃűder, Jan Marcusson, Mony de Leon, Harald Hampel, Philip Scheltens, Tuula PirttilÃd', Anders Wallin, Maria Eriksdotter JÃűnhagen, Lennart Minthon, Bengt Winblad, and Kaj Blennow. Csf biomarkers and incipient alzheimer disease in patients with mild cognitive impairment. *JAMA*, 302(4):385–393, Jul 2009. doi: 10.1001/jama.2009.1064. URL http://dx.doi.org/10.1001/jama.2009.1064.

Jonathan McConathy and Yvette I Sheline. Imaging biomarkers associated with cognitive decline: a review. *Biological psychiatry*, 77(8):685–692, 2015.

G. McKhann, D. Drachman, M. Folstein, R. Katzman, D. Price, and E. M. Stadlan. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology*, 34(7):939–944, July 1984. ISSN 0028-3878. URL http://www.neurology.org/content/34/7/939.abstract.

Guy M McKhann, David S Knopman, Howard Chertkow, Bradley T Hyman, Clifford R Jack, Claudia H Kawas, William E Klunk, Walter J Koroshetz, Jennifer J Manly, Richard Mayeux, et al. The diagnosis of dementia due to alzheimerâĂŹs disease: Recommendations from the national institute on aging-alzheimerâĂŹs association workgroups on diagnostic guidelines for alzheimer's disease. *Alzheimer's & dementia*, 7 (3):263–269, 2011.

H Brendan McMahan, Gary Holt, David Sculley, Michael Young, Dietmar Ebner, Julian Grady, Lan Nie, Todd Phillips, Eugene Davydov, Daniel Golovin, et al. Ad click prediction: a view from the trenches. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1222–1230. ACM, 2013.

V. Menon, K.O. Lim, J.H. Anderson, J. Johnson, and A. Pfefferbaum. Design and efficacy of a head-coil bite bar for reducing movement-related artifacts during functional mri scanning. *Behavior Research Methods, Instruments, & Computers*, 29 (4):589–594, 1997. ISSN 0743-3808. doi: 10.3758/BF03210613. URL http://dx.doi.org/10.3758/BF03210613.

Michal Mikl, Radek Marecek, Petr HlustÃ■k, Martina PavlicovÃą, Ales Drastich, Pavel Chlebus, Milan BrÃązdil, and Petr Krupa. Effects of spatial smoothing on fmri group inferences. *Magnetic resonance imaging*, 26(4):490–503, May 2008. ISSN 0730-725X. URL http://view.ncbi.nlm.nih.gov/pubmed/18060720.

Michael P. Milham, Damien Fair, Maarten Mennes, and Stewart H. Mostofsky. The adhd-200 consortium: a model to advance the translational potential of neuroimaging in clinical neuroscience. *Frontiers in Systems Neuroscience*, 6(62), 2012. ISSN 1662-5137. doi: 10.3389/fnsys.2012.00062. URL http://www.frontiersin.org/systems_neuroscience/10.3389/fnsys.2012.00062/full.

Chandan Misra, Yong Fan, and Christos Davatzikos. Baseline and longitudinal patterns of brain atrophy in mci patients, and their use in prediction of short-term conversion to ad: results from adni. *Neuroimage*, 44(4):1415–1422, 2009.

Elaheh Moradi, Antonietta Pepe, Christian Gaser, Heikki Huttunen, Jussi Tohka, Alzheimer's Disease Neuroimaging Initiative, et al. Machine learning framework for early mri-based alzheimer's conversion prediction in mci subjects. *Neuroimage*, 104: 398–412, 2015.

John C Morris. Early-stage and preclinical alzheimer disease. *Alzheimer Dis Assoc Disord*, 19(3):163–165, 2005.

Janaina Mourao-Miranda, Arun L W Bokde, Christine Born, Harald Hampel, and Martin Stetter. Classifying brain states and determining the discriminating activation patterns: Support vector machine on functional mri data. *Neuroimage*, 28(4):980–995, Dec 2005. doi: 10.1016/j.neuroimage.2005.06.070. URL `http://dx.doi.org/10.1016/j.neuroimage.2005.06.070`.

Susanne G Mueller, Michael W Weiner, Leon J Thal, Ronald C Petersen, Clifford Jack, William Jagust, John Q Trojanowski, Arthur W Toga, and Laurel Beckett. The alzheimer's disease neuroimaging initiative. *Neuroimaging Clin N Am*, 15(4):869–77, xi–xii, Nov 2005. doi: 10.1016/j.nic.2005.09.008. URL `http://dx.doi.org/10.1016/j.nic.2005.09.008`.

Kevin Murphy, Rasmus M. Birn, Daniel A. Handwerker, Tyler B. Jones, and Peter A. Bandettini. The impact of global signal regression on resting state correlations: are anti-correlated networks introduced? *NeuroImage*, 44(3):893–905, February 2009. ISSN 1095-9572. doi: 10.1016/j.neuroimage.2008.09.036. URL `http://dx.doi.org/10.1016/j.neuroimage.2008.09.036`.

Alexandru Niculescu-Mizil and Rich Caruana. Predicting good probabilities with supervised learning. In *Proceedings of the 22nd international conference on Machine learning*, pages 625–632. ACM, 2005.

Jared Nielsen, Brandon Zielinski, Thomas Fletcher, Andrew Alexander, Nicholas Lange, Erin Bigler, Janet Lainhart, and Jeffrey Anderson. Multisite functional connectivity mri classification of autism: Abide results. *Frontiers in human neuroscience*, 7:–,

2013. ISSN 1662-5161. URL http://view.ncbi.nlm.nih.gov/pubmed/24093016.

Ilia Nouretdinov, Sergi Costafreda, Alexander Gammerman, Alexey Chervonenkis, Vladimir Vovk, Vladimir Vapnik, and Cynthia Fu. Machine learning classification with confidence: Application of transductive conformal predictors to mri-based diagnostic and prognostic markers in depression. *NeuroImage*, 56(2):809–813, May 2011. ISSN 10538119. URL http://dx.doi.org/10.1016/j.neuroimage.2010.05.023.

S. Ogawa, T. M. Lee, A. R. Kay, and D. W. Tank. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences*, 87(24):9868–9872, December 1990. ISSN 1091-6490. doi: 10.1073/pnas.87.24.9868. URL http://dx.doi.org/10.1073/pnas.87.24.9868.

Pierre Orban, Cécile Madjar, Mélissa Savard, Christian Dansereau, Angela Tam, Samir Das, Alan C. Evans, Pedro Rosa-Neto, John C.S. Breitner, and Pierre Bellec. Test-retest resting-state fMRI in healthy elderly persons with a family history of alzheimer's disease. *Scientific Data*, 2:150043, oct 2015. doi: 10.1038/sdata.2015.43. URL http://dx.doi.org/10.1038/sdata.2015.43.

Pierre Orban, Christian Dansereau, Laurence Desbois, Violaine Mongeau-Pérusse, Charles-Édouard Giguère, Hien Nguyen, Adrianna Mendrek, Emmanuel Stip, and Pierre Bellec. Multisite generalizability of schizophrenia diagnosis classification based on functional brain connectivity. *Schizophrenia Research*, pages –, 2017a. ISSN 0920-9964. doi: https://doi.org/10.1016/j.schres.2017.05.027. URL http://www.sciencedirect.com/science/article/pii/S092099641730302X.

Pierre Orban, Martin Desseilles, Adrianna Mendrek, Josiane Bourque, Pierre Bellec, and Emmanuel Stip. Altered brain connectivity in patients with schizophrenia is consistent across cognitive contexts. *Journal of psychiatry & neuroscience: JPN*, 42(1):17, 2017b.

Pierre Orban, Angela Tam, Sebastian Urchs, Melissa Savard, Cecile Madjar, Aman-Preet Badhwar, Christian Dansereau, Jacob Vogel, Amir Shmuel, Alain Dagher, Sylvia Villeneuve, Judes Poirier, Pedro Rosa-Neto, John Breitner, Pierre Bellec, , and . Subtypes of functional brain connectivity as early markers of neurodegeneration in alzheimer's disease. *bioRxiv*, 2017c. doi: 10.1101/195164. URL `https://www.biorxiv.org/content/early/2017/09/28/195164`.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

R. C. Petersen. Mild cognitive impairment as a diagnostic entity. *J Intern Med*, 256(3): 183–194, Sep 2004. doi: 10.1111/j.1365-2796.2004.01388.x. URL `http://dx.doi.org/10.1111/j.1365-2796.2004.01388.x`.

William Pettersson-Yeo, Paul Allen, Stefania Benetti, Philip McGuire, and Andrea Mechelli. Dysconnectivity in schizophrenia: where are we now? *Neuroscience & Biobehavioral Reviews*, 35(5):1110–1124, 2011.

Belinda Phipson and Gordon K Smyth. Permutation p-values should never be zero: calculating exact p-values when permutations are randomly drawn. *Statistical applications in genetics and molecular biology*, 9(1), 2010.

John Platt et al. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in large margin classifiers*, 10(3):61–74, 1999.

RA Poldrack, Eliza Congdon, William Triplett, KJ Gorgolewski, KH Karlsgodt, JA Mumford, FW Sabb, NB Freimer, ED London, TD Cannon, et al. A phenome-wide examination of neural and cognitive function. *Scientific data*, 3:160110, 2016.

Jonathan D. Power, Alexander L. Cohen, Steven M. Nelson, Gagan S. Wig, Kelly A. Barnes, Jessica A. Church, Alecia C. Vogel, Timothy O. Laumann, Fran M. Miezin,

Bradley L. Schlaggar, and Steven E. Petersen. Functional Network Organization of the Human Brain. *Neuron*, 72(4):665–678, November 2011. ISSN 08966273. doi: 10.1016/j.neuron.2011.09.006. URL http://dx.doi.org/10.1016/j.neuron.2011.09.006.

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. ISSN 1095-9572. doi: 10.1016/j.neuroimage.2011.10.018. URL http://dx.doi.org/10.1016/j.neuroimage.2011.10.018.

Jonathan D. Power, Anish Mitra, Timothy O. Laumann, Abraham Z. Snyder, Bradley L. Schlaggar, and Steven E. Petersen. Methods to detect, characterize, and remove motion artifact in resting state fMRI. *NeuroImage*, 84:320–341, January 2014. doi: 10.1016/j.neuroimage.2013.08.048. URL http://dx.doi.org/10.1016/j.neuroimage.2013.08.048.

Olivier Querbes, Florent Aubry, Jérémie Pariente, Jean-Albert Lotterie, Jean-François Démonet, Véronique Duret, Michèle Puel, Isabelle Berry, Jean-Claude Fort, Pierre Celsis, et al. Early diagnosis of alzheimer's disease using cortical thickness: impact of cognitive reserve. *Brain*, 132(8):2036–2047, 2009.

Gil D Rabinovici, Maria C Carrillo, Mark Forman, Susan DeSanti, David S Miller, Nicholas Kozauer, Ronald C Petersen, Christopher Randolph, David S Knopman, Eric E Smith, et al. Multiple comorbid neuropathologies in the setting of alzheimer's disease neuropathology and implications for drug development. *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, 3(1):83–91, 2017.

Mehdi Rahim, Bertrand Thirion, Danilo Bzdok, Irène Buvat, and Gaël Varoquaux. Joint prediction of multiple scores captures better individual traits from brain images. *NeuroImage*, 2017.

Saima Rathore, Mohamad Habes, Muhammad Aksam Iftikhar, Amanda Shacklett, and

Christos Davatzikos. A review on neuroimaging-based classification studies and associated feature extraction methods for alzheimer's disease and its prodromal stages. *NeuroImage*, 2017.

Ziad S. Saad, Stephen J. Gotts, Kevin Murphy, Gang Chen, Hang Joon J. Jo, Alex Martin, and Robert W. Cox. Trouble at rest: how correlation patterns and group differences become distorted after global signal regression. *Brain connectivity*, 2(1):25–32, 2012. ISSN 2158-0022. doi: 10.1089/brain.2012.0080. URL http://dx.doi.org/10.1089/brain.2012.0080.

Sohrab Saeb, Luca Lonini, Arun Jayaraman, David C Mohr, and Konrad P Kording. Voodoo machine learning for clinical predictions. *bioRxiv*, page 059774, 2016.

R. Salvador, J. Suckling, M. R. Coleman, J. D. Pickard, D. Menon, and E. Bullmore. Neurophysiological architecture of functional magnetic resonance images of human brain. *Cereb Cortex*, 15(9):1332–1342, September 2005. ISSN 1047-3211. URL http://view.ncbi.nlm.nih.gov/pubmed/15635061.

R. Salvador, A. Martínez, E. Pomarol-Clotet, S. Sarró, J. Suckling, and E. Bullmore. Frequency based mutual information measures between clusters of brain regions in functional magnetic resonance imaging. *NeuroImage*, 35(1):83–88, March 2007. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2006.12.001. URL http://dx.doi.org/10.1016/j.neuroimage.2006.12.001.

Christian Salvatore, Antonio Cerasa, Petronilla Battista, Maria Carla Gilardi, Aldo Quattrone, and Isabella Castiglioni. Magnetic resonance imaging biomarkers for the early diagnosis of alzheimer's disease: a machine learning approach. *Frontiers in neuroscience*, 9:307, 2015.

Leonhard Schilbach, F Hoffstaedter, V Müller, EC Cieslik, R Goya-Maldonado, S Trost, C Sorg, V Riedl, R Jardri, I Sommer, et al. Transdiagnostic commonalities and differences in resting state functional connectivity of the default mode network in schizophrenia and major depression. *NeuroImage: Clinical*, 10:326–335, 2016.

Ganesh M Shankar, Shaomin Li, Tapan H Mehta, Amaya Garcia-Munoz, Nina E Shepardson, Imelda Smith, Francesca M Brett, Michael A Farrell, Michael J Rowan, Cynthia A Lemere, Ciaran M Regan, Dominic M Walsh, Bernardo L Sabatini, and Dennis J Selkoe. Amyloid-beta protein dimers isolated directly from alzheimer's brains impair synaptic plasticity and memory. *Nat Med*, 14(8):837–842, Aug 2008. doi: 10.1038/nm1782. URL http://dx.doi.org/10.1038/nm1782.

Zarrar Shehzad, Clare M. Kelly, Philip T. Reiss, Dylan G. Gee, Kristin Gotimer, Lucina Q. Uddin, Sang Han H. Lee, Daniel S. Margulies, Amy Krain K. Roy, Bharat B. Biswal, Eva Petkova, F. Xavier Castellanos, and Michael P. Milham. The resting brain: unconstrained yet reliable. *Cerebral cortex (New York, N.Y. : 1991)*, 19(10): 2209–2229, October 2009. ISSN 1460-2199. doi: 10.1093/cercor/bhn256. URL http://dx.doi.org/10.1093/cercor/bhn256.

Yvette Sheline and Marcus Raichle. Resting state functional connectivity in preclinical alzheimerâĂŹs disease. *Biological Psychiatry*, 74(5):340–347, September 2013. ISSN 00063223. URL http://dx.doi.org/10.1016/j.biopsych.2012.11.028.

Yvette I Sheline, Joseph L Price, Zhizi Yan, and Mark A Mintun. Resting-state functional mri in depression unmasks increased connectivity between networks via the dorsal nexus. *Proceedings of the National Academy of Sciences*, 107(24):11020–11025, 2010.

Kristina C Skåtun, Tobias Kaufmann, Nhat Trung Doan, Dag Alnæs, Aldo Córdova-Palomera, Erik G Jönsson, Helena Fatouros-Bergman, Lena Flyckt, KaSP, Ingrid Melle, et al. Consistent functional connectivity alterations in schizophrenia spectrum disorder: A multisite study. *Schizophrenia bulletin*, 43(4):914–924, 2016.

John G Sled, Alex P Zijdenbos, and Alan C Evans. A nonparametric method for automatic correction of intensity nonuniformity in mri data. *IEEE transactions on medical imaging*, 17(1):87–97, 1998.

Stephen M. Smith. Fast robust automated brain extraction. *Hum. Brain Mapp.*, 17 (3):143–155, November 2002. ISSN 1065-9471. doi: 10.1002/hbm.10062. URL http://dx.doi.org/10.1002/hbm.10062.

Reisa A. Sperling, Paul S. Aisen, Laurel A. Beckett, David A. Bennett, Suzanne Craft, Anne M. Fagan, Takeshi Iwatsubo, Clifford R. Jack, Jeffrey Kaye, Thomas J. Montine, Denise C. Park, Eric M. Reiman, Christopher C. Rowe, Eric Siemers, Yaakov Stern, Kristine Yaffe, Maria C. Carrillo, Bill Thies, Marcelle Morrison-Bogorad, Molly V. Wagster, and Creighton H. Phelps. Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia*, 7(3):280–292, May 2011. doi: 10.1016/j.jalz.2011.03.003. URL http://dx.doi.org/10.1016/j.jalz.2011.03.003.

Bradley P Sutton, Joshua Goh, Andrew Hebrank, Robert C Welsh, Michael WL Chee, and Denise C Park. Investigation and validation of intersite fmri studies using the same imaging hardware. *Journal of Magnetic Resonance Imaging*, 28(1):21–28, 2008.

Angela Tam, Christian Dansereau, AmanPreet Badhwar, Pierre Orban, Sylvie Belleville, Howard Chertkow, Alain Dagher, Alexandru Hanganu, Oury Monchi, Pedro Rosa-Neto, Amir Shmuel, Seqian Wang, John Breitner, and Pierre Bellec. Common effects of amnestic mild cognitive impairment on resting-state connectivity across four independent studies. *Frontiers in Aging Neuroscience*, 7(242), 2015. ISSN 1663-4365. doi: 10.3389/fnagi.2015.00242. URL http://www.frontiersin.org/aging_neuroscience/10.3389/fnagi.2015.00242/abstract.

Lixia Tian, Tianzi Jiang, Meng Liang, Xiaobo Li, Yong He, Kun Wang, Bingli Cao, and Tao Jiang. Stabilities of negative correlations between blood oxygen level-dependent signals associated with sensory and motor cortices. *Hum Brain Mapp*, 28(7):681–690, Jul 2007. doi: 10.1002/hbm.20300. URL http://dx.doi.org/10.1002/hbm.20300.

Jussi Tohka, Alex Zijdenbos, and Alan Evans. Fast and robust parameter estimation for statistical partial volume models in brain mri. *Neuroimage*, 23(1):84–97, 2004.

Dardo Tomasi and Nora D. Volkow. Functional connectivity density mapping. *Proceedings of the National Academy of Sciences*, 107(21):9885–9890, May 2010. doi: 10.1073/pnas.1001414107. URL http://dx.doi.org/10.1073/pnas.1001414107.

Paule-Joanne Toussaint, Vincent Perlbarg, Pierre Bellec, Serge Desarnaud, Lucette Lacomblez, Julien Doyon, Marie Odile Habert, and Habib Benali. Resting state fdg-pet functional connectivity as an early biomarker of alzheimer's disease using conjoint univariate and independent component analyses. *NeuroImage*, 63 2:936–46, 2012.

Martijn van den Heuvel, Rene Mandl, and Hilleke Hulshoff Pol. Normalized Cut Group Clustering of Resting-State fMRI Data. *PLoS ONE*, 3(4):e2001+, April 2008. doi: 10.1371/journal.pone.0002001. URL http://dx.doi.org/10.1371/journal.pone.0002001.

Koene R. Van Dijk, Trey Hedden, Archana Venkataraman, Karleyton C. Evans, Sara W. Lazar, and Randy L. Buckner. Intrinsic functional connectivity as a tool for human connectomics: theory, properties, and optimization. *Journal of neurophysiology*, 103 (1):297–321, January 2010. ISSN 1522-1598. doi: 10.1152/jn.00783.2009. URL http://dx.doi.org/10.1152/jn.00783.2009.

Koene RA Van Dijk, Mert R Sabuncu, and Randy L Buckner. The influence of head motion on intrinsic functional connectivity mri. *Neuroimage*, 59(1):431–438, 2012.

G. Vanhoutte, M. Verhoye, and A. Van der Linden. Changing body temperature affects the t2* signal in the rat brain and reveals hypothalamic activity. *Magn Reson Med*, 55(5):1006–1012, May 2006. doi: 10.1002/mrm.20861. URL http://dx.doi.org/10.1002/mrm.20861.

Gaël Varoquaux. Cross-validation failure: small sample sizes lead to large error bars. *NeuroImage*, 2017.

Gaël Varoquaux, Pradeep Reddy Raamana, Denis A Engemann, Andrés Hoyos-Idrobo, Yannick Schwartz, and Bertrand Thirion. Assessing and tuning brain decoders: cross-validation, caveats, and guidelines. *NeuroImage*, 145:166–179, 2017.

Anthony B. Waites, Regula S. Briellmann, Michael M. Saling, David F. Abbott, and Graeme D. Jackson. Functional connectivity networks are disrupted in left temporal lobe epilepsy. *Annals of neurology*, 59(2):335–343, February 2006. doi: 10.1002/ana.20733. URL http://dx.doi.org/10.1002/ana.20733.

Stéfan van der Walt, S Chris Colbert, and Gael Varoquaux. The numpy array: a structure for efficient numerical computation. *Computing in Science & Engineering*, 13(2): 22–30, 2011.

Dawei Wang, Bing Liu, Wen Qin, Junping Wang, Yunting Zhang, Tianzi Jiang, and Chunshui Yu. KIBRA gene variants are associated with synchronization within the default-mode and executive control networks. *NeuroImage*, December 2012. ISSN 10538119. doi: 10.1016/j.neuroimage.2012.12.022. URL http://dx.doi.org/10.1016/j.neuroimage.2012.12.022.

Lei Wang, Kathryn I Alpert, Vince D Calhoun, Derin J Cobia, David B Keator, Margaret D King, Alexandr Kogan, Drew Landis, Marcelo Tallis, Matthew D Turner, et al. Schizconnect: Mediating neuroimaging databases on schizophrenia and related disorders for large-scale integration. *NeuroImage*, 124:1155–1167, 2016.

Steve Waterhouse, David MacKay, Tony Robinson, et al. Bayesian methods for mixtures of experts. *Advances in neural information processing systems*, pages 351–357, 1996.

Thomas Wolfers, Jan K Buitelaar, Christian F Beckmann, Barbara Franke, and Andre F Marquand. From estimating activation locality to predicting disorder: a review of pattern recognition for neuroimaging-based psychiatric diagnostics. *Neuroscience & Biobehavioral Reviews*, 57:328–349, 2015.

Choong-Wan Woo, Luke J Chang, Martin A Lindquist, and Tor D Wager. Building better

biomarkers: brain models in translational neuroimaging. *Nature neuroscience*, 20(3): 365–377, 2017.

K. J. Worsley and K. J. Friston. Analysis of fMRI Time-Series Revisited—Again. *NeuroImage*, 2(3):173–181, September 1995. ISSN 10538119. doi: 10.1006/nimg.1995.1023. URL http://dx.doi.org/10.1006/nimg.1995.1023.

Ting-Fan Wu, Chih-Jen Lin, and Ruby C Weng. Probability estimates for multi-class classification by pairwise coupling. *Journal of Machine Learning Research*, 5(Aug): 975–1005, 2004.

Lele Xu, Xia Wu, Kewei Chen, and Li Yao. Multi-modality sparse representation-based classification for alzheimer's disease and mild cognitive impairment. *Computer methods and programs in biomedicine*, 122(2):182–190, 2015.

Chao-Gan Yan, Brian Cheung, Clare Kelly, Stan Colcombe, R. Cameron Craddock, Adriana Di Martino, Qingyang Li, Xi-Nian Zuo, F. Xavier Castellanos, and Michael P. Milham. A comprehensive assessment of regional variation in the impact of head micromovements on functional connectomics. *NeuroImage*, 76:183–201, August 2013a. doi: 10.1016/j.neuroimage.2013.03.004. URL http://dx.doi.org/10.1016/j.neuroimage.2013.03.004.

Chao-Gan G. Yan, Cameron C. Craddock, Xi-Nian N. Zuo, Yu-Feng F. Zang, and Michael P. Milham. Standardizing the intrinsic brain: towards robust measurement of inter-individual variation in 1000 functional connectomes. *NeuroImage*, 80:246–262, October 2013b. URL http://view.ncbi.nlm.nih.gov/pubmed/23631983.

Chaogan Yan, Dongqiang Liu, Yong He, Qihong Zou, Chaozhe Zhu, Xinian Zuo, Xiangyu Long, and Yufeng Zang. Spontaneous Brain Activity in the Default Mode Network Is Sensitive to Different Resting-State Conditions with Limited Cognitive Load. *PLoS ONE*, 4(5):e5743+, May 2009. ISSN 1932-6203. doi: 10.1371/journal.pone.0005743. URL http://dx.doi.org/10.1371/journal.pone.0005743.

Hong Yang, Xiang-Yu Long, Yihong Yang, Hao Yan, Chao-Zhe Zhu, Xiang-Ping Zhou, Yu-Feng Zang, and Qi-Yong Gong. Amplitude of low frequency fluctuation within visual areas revealed by resting-state functional mri. *Neuroimage*, 36(1):144–152, May 2007. doi: 10.1016/j.neuroimage.2007.01.054. URL http://dx.doi.org/10.1016/j.neuroimage.2007.01.054.

B. Thomas Yeo, Fenna M. Krienen, Jorge Sepulcre, Mert R. Sabuncu, Danial Lashkari, Marisa Hollinshead, Joshua L. Roffman, Jordan W. Smoller, Lilla Zöllei, Jonathan R. Polimeni, Bruce Fischl, Hesheng Liu, and Randy L. Buckner. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of neurophysiology*, 106(3):1125–1165, September 2011. ISSN 1522-1598. doi: 10.1152/jn.00338.2011. URL http://dx.doi.org/10.1152/jn.00338.2011.

Zhiwen Yu, Hau-San Wong, and Hongqiang Wang. Graph-based consensus clustering for class discovery from gene expression data. *Bioinformatics*, 23(21):2888–2896, November 2007. doi: 10.1093/bioinformatics/btm463. URL http://dx.doi.org/10.1093/bioinformatics/btm463.

Y. Yuan, Z-X. Gu, and W-S. Wei. Fluorodeoxyglucose-positron-emission tomography, single-photon emission tomography, and structural mr imaging for prediction of rapid conversion to alzheimer disease in patients with mild cognitive impairment: a meta-analysis. *AJNR Am J Neuroradiol*, 30(2):404–410, Feb 2009. doi: 10.3174/ajnr.A1357. URL http://dx.doi.org/10.3174/ajnr.A1357.

Yu-Feng F. Zang, Yong He, Chao-Zhe Z. Zhu, Qing-Jiu J. Cao, Man-Qiu Q. Sui, Meng Liang, Li-Xia X. Tian, Tian-Zi Z. Jiang, and Yu-Feng F. Wang. Altered baseline brain activity in children with ADHD revealed by resting-state functional MRI. *Brain & development*, 29(2):83–91, March 2007. ISSN 0387-7604. doi: 10.1016/j.braindev.2006.07.002. URL http://dx.doi.org/10.1016/j.braindev.2006.07.002.

Xiuming Zhang, Elizabeth C Mormino, Nanbo Sun, Reisa A Sperling, Mert R Sabuncu, BT Thomas Yeo, Michael W Weiner, Paul Aisen, Michael Weiner, Ronald Petersen,

et al. Bayesian model reveals latent atrophy factors with dissociable cognitive trajectories in alzheimerâĂŹs disease. *Proceedings of the National Academy of Sciences*, 113(42):E6535–E6544, 2016.

Weihao Zheng, Zhijun Yao, Bin Hu, Xiang Gao, Hanshu Cai, and Philip Moore. Novel cortical thickness pattern for accurate detection of alzheimerâĂŹs disease. *Journal of Alzheimer's Disease*, 48(4):995–1008, 2015.

Yuan Zhou, Meng Liang, Lixia Tian, Kun Wang, Yihui Hao, Haihong Liu, Zhening Liu, and Tianzi Jiang. Functional disintegration in paranoid schizophrenia using resting-state fmri. *Schizophr Res*, 97(1-3):194–205, Dec 2007. doi: 10.1016/j.schres.2007.05.029. URL http://dx.doi.org/10.1016/j.schres.2007.05.029.

Yuan Zhou, Ni Shu, Yong Liu, Ming Song, Yihui Hao, Haihong Liu, Chunshui Yu, Zhening Liu, and Tianzi Jiang. Altered resting-state functional connectivity and anatomical connectivity of hippocampus in schizophrenia. *Schizophr Res*, 100(1-3):120–132, Mar 2008. doi: 10.1016/j.schres.2007.11.039. URL http://dx.doi.org/10.1016/j.schres.2007.11.039.

Xiaofeng Zhu, Heung-Il Suk, and Dinggang Shen. A novel matrix-similarity based loss function for joint regression and classification in ad diagnosis. *NeuroImage*, 100:91–105, 2014.

Chen Zu, Biao Jie, Mingxia Liu, Songcan Chen, Dinggang Shen, Daoqiang Zhang, AlzheimerâĂŹs Disease Neuroimaging Initiative, et al. Label-aligned multi-task feature learning for multimodal classification of alzheimerâĂŹs disease and mild cognitive impairment. *Brain imaging and behavior*, 10(4):1148–1159, 2016.

Xi-Nian Zuo, Ting Xu, Lili Jiang, Zhi Yang, Xiao-Yan Cao, Yong He, Yu-Feng Zang, F. Xavier Castellanos, and Michael P. Milham. Toward reliable characterization of functional homogeneity in the human brain: Preprocessing, scan duration, imaging resolution and computational space. *NeuroImage*, October 2012. ISSN 10538119. doi:

10.1016/j.neuroimage.2012.10.017. URL http://dx.doi.org/10.1016/j.neuroimage.2012.10.017.

Xi-Nian Zuo, Jeffrey S Anderson, Pierre Bellec, Rasmus M Birn, Bharat B Biswal, Janusch Blautzik, John C S Breitner, Randy L Buckner, Vince D Calhoun, F Xavier Castellanos, Antao Chen, Bing Chen, Jiangtao Chen, Xu Chen, Stanley J Colcombe, William Courtney, R Cameron Craddock, Adriana Di Martino, Hao-Ming Dong, Xiaolan Fu, Qiyong Gong, Krzysztof J Gorgolewski, Ying Han, Ye He, Yong He, Erica Ho, Avram Holmes, Xiao-Hui Hou, Jeremy Huckins, Tianzi Jiang, Yi Jiang, William Kelley, Clare Kelly, Margaret King, Stephen M LaConte, Janet E Lainhart, Xu Lei, Hui-Jie Li, Kaiming Li, Kuncheng Li, Qixiang Lin, Dongqiang Liu, Jia Liu, Xun Liu, Yijun Liu, Guangming Lu, Jie Lu, Beatriz Luna, Jing Luo, Daniel Lurie, Ying Mao, Daniel S Margulies, Andrew R Mayer, Thomas Meindl, Mary E Meyerand, Weizhi Nan, Jared A Nielsen, David O'Connor, David Paulsen, Vivek Prabhakaran, Zhigang Qi, Jiang Qiu, Chunhong Shao, Zarrar Shehzad, Weijun Tang, Arno Villringer, Huiling Wang, Kai Wang, Dongtao Wei, Gao-Xia Wei, Xu-Chu Weng, Xuehai Wu, Ting Xu, Ning Yang, Zhi Yang, Yu-Feng Zang, Lei Zhang, Qinglin Zhang, Zhe Zhang, Zhiqiang Zhang, Ke Zhao, Zonglei Zhen, Yuan Zhou, Xing-Ting Zhu, and Michael P Milham. An open science resource for establishing reliability and reproducibility in functional connectomics. *Sci Data*, 1:140049, 9 December 2014.

Appendix I

First Appendix

frontiers in Aging Neuroscience

ORIGINAL RESEARCH
published: 24 December 2015
doi: 10.3389/fnagi.2015.00242

Common Effects of Amnestic Mild Cognitive Impairment on Resting-State Connectivity Across Four Independent Studies

Angela Tam[1, 2, 3], Christian Dansereau[3, 4], AmanPreet Badhwar[3, 4], Pierre Orban[2, 3], Sylvie Belleville[3, 4], Howard Chertkow[1], Alain Dagher[1], Alexandru Hanganu[3, 5, 6], Oury Monchi[3, 4, 5, 6], Pedro Rosa-Neto[1, 2], Amir Shmuel[1], Seqian Wang[1, 2], John Breitner[1, 2] and Pierre Bellec[3, 4]* for the Alzheimer's Disease Neuroimaging Initiative[†]*

[1] McGill University, Montreal, QC, Canada, [2] Douglas Mental Health University Institute, Research Centre, Montreal, QC, Canada, [3] Centre de Recherche de L'institut Universitaire de Gériatrie de Montréal, Montreal, QC, Canada, [4] Université de Montréal, Montreal, QC, Canada, [5] University of Calgary, Calgary, AB, Canada, [6] Hotchkiss Brain Institute, Calgary, AB, Canada

Resting-state functional connectivity is a promising biomarker for Alzheimer's disease. However, previous resting-state functional magnetic resonance imaging studies in Alzheimer's disease and amnestic mild cognitive impairment (aMCI) have shown limited reproducibility as they have had small sample sizes and substantial variation in study protocol. We sought to identify functional brain networks and connections that could consistently discriminate normal aging from aMCI despite variations in scanner manufacturer, imaging protocol, and diagnostic procedure. We therefore combined four datasets collected independently, including 112 healthy controls and 143 patients with aMCI. We systematically tested multiple brain connections for associations with aMCI using a weighted average routinely used in meta-analyses. The largest effects involved the superior medial frontal cortex (including the anterior cingulate), dorsomedial prefrontal cortex, striatum, and middle temporal lobe. Compared with controls, patients with aMCI exhibited significantly decreased connectivity between default mode network nodes and between regions of the cortico-striatal-thalamic loop. Despite the heterogeneity of methods among the four datasets, we identified common aMCI-related connectivity changes with small to medium effect sizes and sample size estimates recommending a minimum of 140 to upwards of 600 total subjects to achieve adequate statistical power in the context of a multisite study with 5–10 scanning sites and about 10 subjects per group and per site. If our findings can be replicated and associated with other established biomarkers of Alzheimer's disease (e.g., amyloid and tau quantification), then these functional connections may be promising candidate biomarkers for Alzheimer's disease.

Keywords: fMRI, mild cognitive impairment, connectome, resting-state, default mode network, meta-analysis

OPEN ACCESS

Edited by:
Junfeng Sun,
Shanghai Jiao Tong University, China

Reviewed by:
Donald G. McLaren,
University of Wisconsin-Madison, USA
Rui Li,
Chinese Academy of Sciences, China

****Correspondence:***
Angela Tam
angela.tam@mail.mcgill.ca;
Pierre Bellec
pierre.bellec@criugm.qc.ca

[†]Data used in preparing this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but most of them did not participate in this analysis or writing this report. A complete list of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

Received: *28 August 2015*
Accepted: *10 December 2015*
Published: *24 December 2015*

Citation:
Tam A, Dansereau C, Badhwar A, Orban P, Belleville S, Chertkow H, Dagher A, Hanganu A, Monchi O, Rosa-Neto P, Shmuel A, Wang S, Breitner J, Bellec P for the Alzheimer's Disease Neuroimaging Initiative (2015) Common Effects of Amnestic Mild Cognitive Impairment on Resting-State Connectivity Across Four Independent Studies. Front. Aging Neurosci. 7:242. doi: 10.3389/fnagi.2015.00242

INTRODUCTION

Resting-state connectivity in functional magnetic resonance imaging (fMRI) captures the spatial coherence of spontaneous fluctuations in blood oxygenation. Resting-state fMRI is a promising technique that may be useful as an early biomarker for Alzheimer's disease (AD), a neurodegenerative process that develops over decades before patients suffer from dementia. The possibility that disturbed resting-state connectivity may be an early marker for AD is supported by studies of mild cognitive impairment (MCI), a disorder characterized by objective cognitive deficits without dementia, i.e., without impairment in activities of daily living, and more specifically by studies of amnestic MCI (aMCI), the most common subtype of MCI characterized by memory deficits (Petersen et al., 2001). These studies showed altered functional connectivity in MCI compared with cognitively normal elderly (CN; Sorg et al., 2007; Bai et al., 2009; Liang et al., 2012; Wu et al., 2014), but they relied on small sample sizes ($n = \sim 40$) and differed in many aspects of their protocols, e.g., recruitment and image acquisition procedures. If resting-state fMRI is to serve as a useful biomarker of AD, or any pathology, for clinical practice or research, we must determine if changes in functional connectivity differences between groups of subjects are robust to such variation in study protocols. Therefore, we sought to identify brain connections that showed consistent MCI-related changes across multiple independent studies. If such connections exist, they may be used as targets to be examined alongside other established AD biomarkers (e.g., amyloid and tau measures) in order to validate resting-state fMRI's potential as a biomarker for AD.

Resting-state connectivity studies have consistently found decreased connectivity between nodes within the default mode network (DMN) in patients with AD or MCI compared with CN (Sorg et al., 2007; Bai et al., 2009; Zhang et al., 2010; Koch et al., 2012; Liang et al., 2012). Less consistent are reports of alterations in the executive attentional, frontoparietal, and anterior temporal networks (Sorg et al., 2007; Zhang et al., 2010; Gour et al., 2011; Agosta et al., 2012; Liang et al., 2012; Wu et al., 2014) due to the literature's bias toward investigating the DMN. Further inconsistencies can be found in some studies that have reported increased connectivity between the middle temporal lobe and other DMN areas in MCI (Qi et al., 2010), while others have reported decreased connectivity between these same regions (Bai et al., 2009) and others have reported no significant differences between MCI and CN (Koch et al., 2012).

One obvious explanation for such inconsistency may be these studies' small sample sizes resulting in low statistical power (Kelly et al., 2012). Beyond this, however, there are other methodological differences that may compromise the comparison of results across independent studies. For example, the criteria for recruiting subjects with MCI, e.g., Petersen (2004) vs. NIA-AA recommendations (Albert et al., 2011) may differ among studies. Different study samples may also reflect different socio-cultural characteristics of recruiting sites, e.g., ethnicity, language, diet, socioeconomic status. The fMRI measurements themselves can also be affected by differences in details of the image acquisition such as scanner make and model (Friedman et al., 2006), sequence parameters such as repetition time, flip angle, or acquisition volume (Friedman and Glover, 2006), experimental design such as eyes-open/eyes-closed (Yan et al., 2009) or experiment duration (van Dijk et al., 2010), and scanning environment such as sound attenuation measures (Elliott et al., 1999), room temperature (Vanhoutte et al., 2006), or head-motion restraint techniques (Edward et al., 2000).

To identify robust changes in resting-state connectivity between aMCI and CN, we implemented a meta-analysis of four independent resting-state fMRI datasets (ADNI2 and three small single-site studies) using a weighted average implemented by Willer et al. (2010). Rather than relying on a priori target regions or connections, we leveraged the large sample size to perform a systematic search of brain connections affected by aMCI, an approach termed a "connectome-wide association study" (Shehzad et al., 2014). In addition, we relied on functionally-defined brain parcellations using an automated clustering procedure and we explored the impact of the number of brain clusters (called resolution) on observed differences (Bellec et al., 2015).

METHODS

Participants

We combined data from four independent studies: the Alzheimer's Disease Neuroimaging Initiative 2 (ADNI2) sample, two samples from the Centre de recherche de l'institut universitaire de gériatrie de Montréal (CRIUGMa and CRIUGMb), and a sample from the Montreal Neurological Institute (MNI; Wu et al., 2014). All participants gave their written informed consent to engage in these studies, which were approved by the research ethics board of the respective institutions, and included consent for data sharing with collaborators as well as secondary analysis. Ethical approval was also obtained at the site of secondary analysis (CRIUGM).

The ADNI2 data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). ADNI was launched in 2003 by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, the Food and Drug Administration, private pharmaceutical companies and non-profit organizations, as a $60 million, 5-year public-private partnership representing efforts of co-investigators from numerous academic institutions and private corporations. ADNI was followed by ADNI-GO and ADNI-2 that included newer techniques. Subjects included in this study were recruited by ADNI-2 from all 13 sites that acquired resting-state fMRI on Philips scanners across North America. For up-to-date information, see www.adni-info.org.

The combined sample included 112 CN and 143 aMCI prior to quality control. After quality control, 99 CN and 129 aMCI remained. In the CN group, the mean age was 72.0 ($s.d. = 7.0$) years, and 37% were men. Mean age of the aMCI subjects was 72.3 ($s.d. = 7.6$) years, and 50% were men. An independent samples t-test did not reveal any significant difference in age between the groups ($t = 0.759, p = 0.448$). A chi-squared test revealed a trend

toward a significant difference in gender distribution between the groups ($\chi^2 = 3.627$, $p = 0.057$). Note that both age and gender were entered as confounding variables in the statistical analysis below. See **Table 1** for sample size and demographic information from the individual studies after passing quality control (for information about the original cohorts before quality control, see Supplementary Table 1).

All subjects underwent cognitive testing (e.g., memory, language, and executive function; see **Table 2** for a list of specific tests used in each study). Exclusion criteria common to all studies included: Contraindications to MRI, presence or history of axis I psychiatric disorders (e.g., depression, bipolar disorder, schizophrenia), presence or history of neurologic disease with potential impact on cognition (e.g., Parkinson's disease), and presence or history of substance abuse. CN subjects could not meet criteria for MCI or dementia. Those with aMCI had memory complaints, objective cognitive loss (based on neuropsychological testing), but had intact functional abilities and did not meet criteria for dementia. In ADNI2, the diagnosis of aMCI was made based on an education adjusted abnormal score on the Logical Memory II subscale (Delayed Paragraph Recall, Paragraph A only) from the Wechsler Memory Scale and a Clinical Dementia Rating (CDR) of 0.5. In both CRIUGMa and CRIUGMb, the diagnosis of aMCI was made based on scores equal to or >1.5 standard deviations below the mean adjusted for age and education on memory tests. At the MNI, the diagnosis of aMCI relied on the Petersen criteria (2004). At both CRIUGMb and MNI, aMCI diagnoses were made with input from a neurologist. See the Supplementary Methods (Datasheet 1 in Supplementary Material) for greater details for each study.

Imaging Data Acquisition

All resting-state fMRI and structural scans were acquired on 3T scanners. We performed analyses on the first usable scan (typically the baseline scan) from ADNI2 and applied clinical diagnoses from the same study time point as the first usable scan for each participant in that dataset. See **Table 3** for acquisition parameters for each sample.

Computational Environment

All experiments were performed using the NeuroImaging Analysis Kit (NIAK[1]; Bellec et al., 2011) version 0.12.18, under CentOS version 6.3 with Octave[2] version 3.8.1 and the Minc toolkit[3] version 0.3.18. Analyses were executed in parallel on the "Guillimin" supercomputer[4], using the pipeline system for Octave and Matlab (Bellec et al., 2012), version 1.0.2. The scripts used for processing can be found on Github[5].

Pre-processing

Each fMRI dataset was corrected for slice timing; a rigid-body motion was then estimated for each time frame, both within and between runs, as well as between one fMRI run and the T1 scan for each subject (Collins and Evans, 1997). The T1 scan was itself non-linearly co-registered to the Montreal Neurological Institute (MNI) ICBM152 stereotaxic symmetric template (Fonov et al., 2011), 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 frames showing a displacement >0.5 mm were removed (Power et al., 2012). A minimum of 50 unscrubbed volumes per run was required for further analysis (13 CN and 14 aMCI were rejected from the original cohort of 112 CN and 143 aMCI). Neither the rate of rejection nor the frame displacement values (before and after scrubbing) varied significantly among the

[1]http://simexp.github.io/niak/.
[2]https://www.gnu.org/software/octave/.
[3]http://www.bic.mni.mcgill.ca/ServicesSoftware/ServicesSoftwareMincToolKit.
[4]http://www.calculquebec.ca/en/resources/compute-servers/guillimin.
[5]https://github.com/SIMEXP/mcinet.

TABLE 1 | Demographic information in all studies after quality control.

		ADNI2	CRIUGMa	CRIUGMb	MNI	Combined sample
CN	**N**	**49**	**18**	**17**	**15**	**99**
	Mean age (*s.d.*)	74.4 (6.8)	71.2 (8.0)	70.4 (4.6)	67.0 (5.7)	72.0 (7.0)
	Number male (%)	21 (43%)	7 (39%)	2 (12%)	7 (47%)	37 (37%)
	Mean years of education (*s.d.*)[a]	16.9 (2.2)	14.9 (2.3)	15.1 (2.8)	15.0 (3.1)	16.0 (2.6)
	MMSE mean (range)	28.7 (25–30)	28.8 (27–30)	n/a	29.0 (27–30)	n/a
	MoCA mean (range)	n/a	27.8 (22–30)	28.4 (26–30)	n/a	n/a
aMCI	**N**	**82**	**8**	**21**	**18**	**129**
	Mean age (*s.d.*)	71.2 (7.3)	79.9 (6.1)	74.8 (7.0)	71.2 (8.1)	72.3 (7.6)
	Number male (%)	43 (52%)	3 (38%)	12 (57%)	7 (39%)	65 (50%)
	Mean years of education (*s.d.*)[a]	16.2 (2.6)	13.7 (3.8)	14.8 (4.2)	13.1 (3.1)	15.5 (3.2)
	MMSE mean (range)	28.1 (24–30)*	26.1 (22–29)*	n/a	26.1 (22–30)*	n/a
	MoCA mean (range)	n/a	23.3 (20–29)*	24.6 (16–29)*	n/a	n/a

MMSE, Mini-mental state examination; MoCA, Montreal Cognitive Assessment.
**Significant difference between aMCI and CN (within study) for independent samples t-test at $p \leq 0.05$.*
[a]Missing values for education for subjects in ADNI2 (1 CN, 1 aMCI), CRIUGMb (2 aMCI), and MNI (3 CN, 6 aMCI).

TABLE 2 | Neuropsychological tests that were used in each study.

Test	ADNI2	CRIUGMa	CRIUGMb	MNI
Mini-mental state examination (MMSE)	x	x		x
Montreal Cognitive Assessment (MoCA)	x	x	x	
Clinical Dementia Rating (CDR)	x		x	
ADAS-Cog	x			
Everyday Cognition (ECog)	x			
Trail making	x (Trails A and B)	x (Trails A and B)	x (Trails A and B)	x (DKEFS)
Boston naming test	x	x	x	x
Digit span		x	x	x
Color-word interference (DKEFS)		x	x	x
Rey auditory verbal learning test	x	x		x
Verbal fluency	x	x	x (MEC)	x (DKEFS)
Clock drawing	x	x		
Visual object and space perception battery		x		
Brixton spatial anticipation test			x	
Hooper visual organization test			x	
Rey complex figure		x	x	x
Aggie figures learning test				x
16-Item free and cued recall (RL/RI-16)			x	
Pyramid and palm trees test		x		
Weschler memory scale—logical memory subtest	x	x	x	

MEC, Montréal évaluation de la communication; DKEFS, Delis–Kaplan Executive Function System.

four samples or between CN and aMCI. The following nuisance covariates were regressed out from fMRI 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 3–10 principal components (median numbers for ADNI2, CRIUGMa, CRIUGMb, and MNI were 9, 6, 7, and 7, respectively, and accounting for 95% variance) of the six rigid-body motion parameters and their squares (Lund et al., 2006; Giove et al., 2009). The fMRI volumes were finally spatially smoothed with a 6 mm isotropic Gaussian blurring kernel. A more detailed description of the pipeline can be found on the NIAK website[6] and Github[7].

Bootstrap Analysis of Stable Clusters (BASC)

We applied a BASC to identify clusters that consistently exhibited similar spontaneous BOLD fluctuations in individual subjects, and were spatially stable across subjects. We first applied a region-growing algorithm to reduce each fMRI dataset into a time × space array, with 957 regions (Bellec et al., 2006). BASC replicates a hierarchical Ward clustering 1000 times and computes the probability that a pair of regions fall in the same cluster, a measure called stability. The region × region stability matrix is fed into a clustering procedure to derive consensus clusters, which are composed of regions with a high average probability of being assigned to the same cluster across all replications. At the individual level, the clustering was applied to the similarity of regional time series, which was replicated using a circular block bootstrap. Consensus clustering was applied to the average individual stability matrix to identify group clusters. The group clustering was replicated via bootstrapping of subjects in the group. A consensus clustering was finally applied on the group stability matrix to generate group consensus clusters.

The cluster procedure was carried out at a specific number of clusters (called resolution). Using a "multiscale stepwise selection" (MSTEPS) method (Bellec, 2013), we determined a subset of resolutions that provided an accurate summary of the group stability matrices generated over a fine grid of resolutions: 4, 6, 12, 22, 33, 65, 111, and 208.

Derivation of Functional Connectomes

For each resolution K, and each pair of distinct clusters, the between-clusters connectivity was measured by the Fisher transform of the Pearson's correlation between the average time series of the clusters. The within-cluster connectivity was the Fisher transform of the average correlation between time series inside the cluster. An individual connectome was thus a $K \times K$ matrix. See **Figures 1A,B** for an illustration of a parcellation and associated connectome.

Statistical Testing

To test for differences between aMCI and CN at a given resolution, we used a general linear model (GLM) for each connection between two clusters. The GLM included an intercept, the age and sex of participants, and the average frame

[6]http://niak.simexp-lab.org/pipe_preprocessing.html.
[7]https://github.com/SIMEXP/mcinet/tree/master/preprocess.

TABLE 3 | Structural and functional scan acquisition parameters.

	ADNI2[a]	CRIUGMa	CRIUGMb	MNI
Scanner manufacturer	**Philips**	**Siemens**	**Siemens**	**Siemens**
STRUCTURAL				
No. channels	8	32	32	32
No. slices	170	176	176	176
Slice thickness (mm)	1.2	1	1	1
In-plane resolution (mm × mm)	1 × 1	1 × 1	1 ×1	1 × 1
Matrix size	256 × 256	240 × 256	256 × 256	256 × 256
FOV (mm^2)	256	240/256	256	256
TR (s)	6.8	2.3	2.53	2.3
TE (ms)	3.09	2.91	1.64	2.98
TI (s)	n/a	0.9	1.2	0.9
FA (°)	9	9	7	9
Slice gap	0	0	0	0
Imaging plane	Sagittal	Sagittal	Sagittal	Sagittal
NEX	1	1	1	1
FUNCTIONAL				
No. runs	1	1	3	3
No. channels	8	32	32	32
No. volumes	140	240	150	160
No. slices	48	33	42	38
Slice thickness (mm)	3.3	4	3.4	3.6
In-plane resolution (mm × mm)	3.3 × 3.3	3 × 3	3.4 × 3.4	3.6 × 3.6
Matrix size	64 × 64	64 × 64	64 × 64	64 × 64
FOV (mm^2)	212	192	218	230
TR (s)	3	2	2.6	2
TE (ms)	30	30	30	30
FA (°)	80	90	90	90
Slice gap	0	0	0	0
Imaging plane	Axial	Axial	Axial	Axial
NEX	1	1	1	1
Total scan time (min:s)	7:00	8:00	19:30	16:00

[a] *http://adni.loni.usc.edu/wp-content/uploads/2011/04/ADNI_3T_Philips_2.6.pdf.*

displacement of the runs involved in the analysis. The contrast of interest (aMCI–CN) was represented by a dummy covariate coding the difference in average connectivity between the two groups. All covariates except the intercept were corrected to a zero mean (**Figure 1C**). The GLM was estimated independently for each scanning protocol. In addition to distinguishing between CRIUGMa, CRIUGMb, MNI, and ADNI2, ADNI2 was subdivided into five sub-studies based on the use of different Philips scanner models (i.e., Achieva, Gemini, Ingenia, Ingenuity, and Intera). We dropped all subjects scanned with Ingenuity (2 CN, 1 aMCI) due to the elimination of all aMCI subjects within that site by the scrubbing procedure and its small sample size. We therefore estimated seven independent GLMs for each protocol (ADNI2-Achieva, ADNI2-Gemini, ADNI2-Ingenia, ADNI2-Intera, CRIUGMa, CRIUGMb, MNI). The estimated effects were combined across all protocols through inverse variance based weighted averaging (Willer et al., 2010; **Figure 1D**).

Resolutions containing fewer than 50 clusters have been suggested to have higher sensitivity based on prior independent work (Bellec et al., 2015). The GLM was first applied at an a priori resolution of $K = 33$, which was the lowest number of clusters for which the DMN could be clearly decomposed into subnetworks (Supplementary Figure 1, visit Figshare for 3D volumes of brain parcellations[8] and see Supplementary Table 2 for a list of the 33 clusters and their numerical IDs). The false-discovery rate (FDR) across connections was controlled at $q^{FDR} \leq 0.1$ (Benjamini and Hochberg, 1995). In addition to the analysis at resolution 33, we assessed the impact of that parameter by replicating the GLM analysis at the seven resolutions selected by MSTEPS (Supplementary Figure 2). We implemented an omnibus test (family-wise error rate $\alpha \leq 0.05$) to assess the overall presence of significant differences between groups, pooling FDR results across all resolutions (Bellec et al., 2015). If the omnibus test across resolutions was not significant, then no test would be deemed significant. Since this omnibus test was significant, we used the FDR threshold of $q \leq 0.1$ to explore single resolutions.

RESULTS

Functional Connectivity Differences Between aMCI and CN

The omnibus test pooling significant differences in connectivity between aMCI and CN across all resolutions was significant at $\alpha \leq 0.05$ ($p \leq 0.0056$). In line with prior observations on independent datasets (Bellec et al., 2015), resolutions containing fewer than 50 clusters were associated with a higher rate of discovery (**Figure 2**). At resolution 33, significant group differences between aMCI and CN were seen across the whole brain (**Figure 3A**). Four brain clusters were associated with 47% of all significant changes found across the connectome: the superior medial frontal cortex (including anterior cingulate), dorsomedial prefrontal cortex, striatum, and middle temporal lobe (**Figures 3B,C**, Supplementary Table 3). Supplementary Table 3 contains a list of parcels that account for all non-redundant significant connectivity differences between aMCI and CN. For example, the first-ranked seed (superior medial frontal cortex) was associated with 13.4% of connections that differ between the groups. The second-ranked seed (dorsomedial prefrontal cortex) was associated with an additional 12.7% of connectivity differences that did not overlap with or were not previously accounted for by the first seed. Note that if a given parcel was associated with a significant effect with another region that ranked in the table, then that parcel may not be listed in the table (i.e., this table is not a comprehensive list of parcels that show significant effects, as a given parcel may involve a region in the table at a higher rank which already accounted for its effects). Given that the top four clusters explained nearly half of the findings, they were further characterized in seed-based connectivity analyses, which revealed that aMCI showed decreased connectivity between DMN nodes and between areas of the cortico-striatal-thalamic loop (**Figure 4**). More specifically, in aMCI compared to CN, the superior medial frontal cortex displayed significantly reduced connectivity with

[8] http://dx.doi.org/10.6084/m9.figshare.1480461.

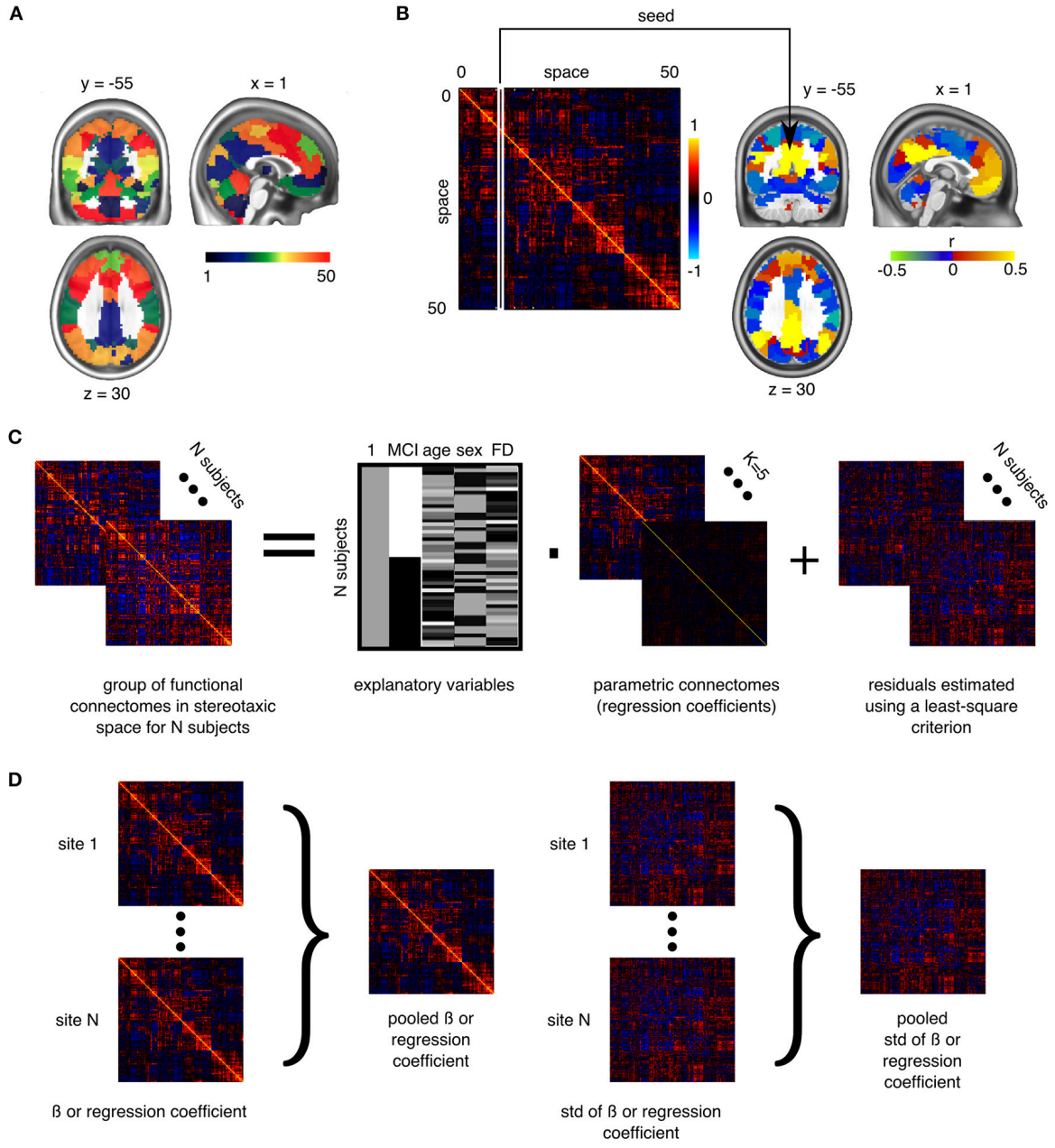

FIGURE 1 | Application of general linear models to connectomes. (A) The brain is functionally parcellated into *K* (e.g., 50) clusters generated through a clustering algorithm. **(B)** The connectome is a $K \times K$ matrix measuring functional connectivity between and within clusters. **(C)** A general linear model is used to test the association between phenotypes and connectomes, independently at each connection, at the group level. **(D)** In a multisite situation, independent site-specific effects are estimated and then pooled through weighted averaging (Willer et al., 2010).

the ventromedial prefrontal cortex, striatum, thalamus, temporal lobes, hippocampus, inferior parietal lobes, and precuneus (**Figure 4A**). aMCI showed reduced connectivity between the dorsomedial prefrontal cortex with temporal lobe regions, ventral frontal areas, thalamus, striatum, and the cuneus (**Figure 4B**). The striatum in aMCI also exhibited decreased connectivity with the sensorimotor cortex, thalamus, and frontal and parietal regions (**Figure 4C**). Lastly, in aMCI, the middle temporal lobe displayed significantly decreased connectivity with the posterior cingulate, precuneus, inferior parietal lobes, hippocampus, and frontal areas (**Figure 4D**).

Sample-Specific Effects

The statistical model we used to combine GLM analyses across sites was based on a weighted average. The possibility thus existed that an effect would be significant in the pooled analysis because it was driven by a very strong effect in a single sample, instead of being consistent across all samples. When we examined

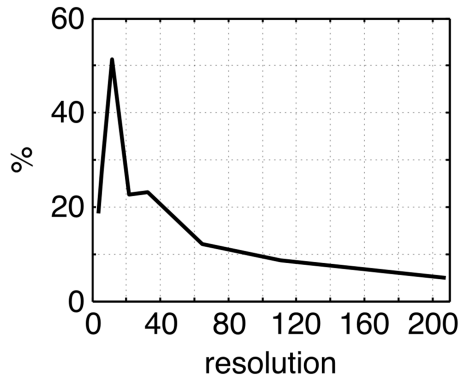

FIGURE 2 | Plot of the percentage of connections identified as significant by the statistical comparison between aMCI and CN across the connectome ($q^{FDR} \leq 0.1$), as a function of the resolutions selected by MSTEPS.

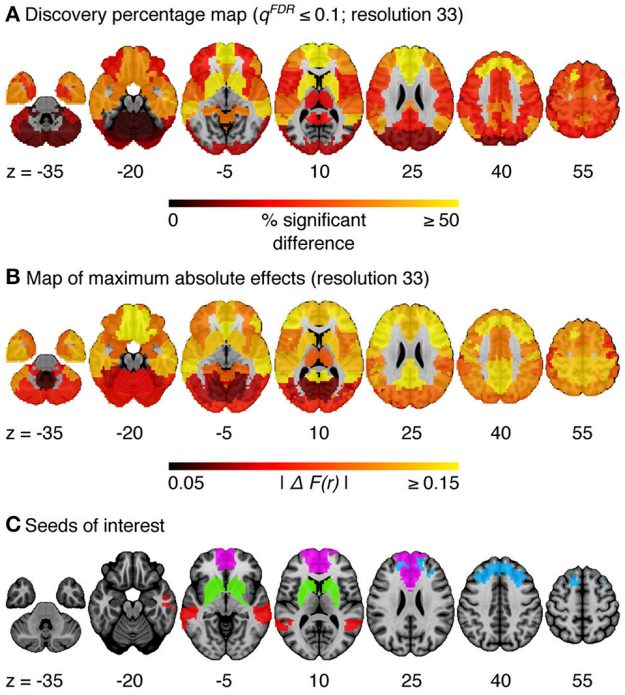

FIGURE 3 | (A) Map of the percentage of connections associated with a given cluster and identified as significant by the statistical comparison between aMCI and CN, at a resolution of 33 clusters ($q^{FDR} \leq 0.1$). **(B)** Maximum absolute difference in average connectivity between aMCI and CN, across all connections associated with a cluster, at resolution 33. ΔF(r) signifies the difference in Fisher-transformed correlation values between the groups. **(C)** Four clusters of interest (superior medial frontal cortex, dorsomedial prefrontal cortex, striatum, middle temporal lobe) were selected out of 33 for further characterization.

effects in each sample independently, we detected no findings or very few significant findings. We then explored the whole brain connectivity of the top four seed regions (superior medial frontal cortex, dorsomedial prefrontal cortex, striatum, and middle temporal lobe) within each sample. The majority of effects found at each sample did not appear to be consistent or reproducible across studies as the comparison between aMCI and CN varied substantially among the seven samples (**Figure 5**, Supplementary Figures 3–5). We assessed the extent at which findings among the seven samples were similar by calculating correlation coefficients across the spatial maps for the average connectivity values in CN, the average connectivity values in aMCI, and differences in connectivity values between aMCI and CN among the samples. We found that the difference maps, contrasting aMCI and CN, were weakly correlated on average across studies and protocols (mean $r = 0.06$, min $r = -0.64$, max $r = 0.69$). The average connectivity maps among studies in both CN and aMCI were generally highly correlated with each other (for CN, mean $r = 0.68$, min $r = -0.16$, max $r = 0.95$; for aMCI, mean $r = 0.67$, min $r = -0.10$, max $r = 0.97$). These results were expected given the small sample sizes of most independent samples (Kelly et al., 2012), but still sobering as the majority of the literature on aMCI and fMRI has used small sample sizes.

However, despite the large observed variations in the spatial distribution of aMCI vs. CN contrasts, there were still clear consistent trends across studies and protocols. We indeed found that aMCI-related connectivity changes that surpassed the FDR threshold in the pooled analysis showed similar trends in the vast majority of samples across seeds and connections, where the independent aMCI samples consistently exhibited decreased connectivity compared to the CN samples (**Figures 5**, **6**, Supplementary Figures 3–5). For example, the pooled analysis revealed that, compared to CN, aMCI exhibited significantly reduced connectivity between the superior medial frontal cortex cortex (the region in which connectivity was most affected by aMCI) and the middle temporal lobes. This change appeared to be common to the majority of the independent samples (**Figures 5**, **6A**). For this particular seed, the change in connectivity was mainly due to regions with positive correlations in CN having smaller correlation values closer to zero in aMCI in the individual samples (**Figures 5**, **6A**). For sample-specific effects in other seeds and connections, please see Supplementary Figures 3–9.

Effect Sizes and Sample Size Estimates

We measured the effect sizes of the difference between groups at each significant connection by calculating Cohen's *d*, via a weighted average of the effect sizes per individual sample. We found small to medium effect sizes, ranging from $d = 0.10$–0.48, with an average effect size of $d = 0.32$. Note that these effect sizes are potentially inflated since we have focussed on significant results only. We also calculated the sample sizes required to achieve 80% power, based on the effect sizes estimated by Cohen's *d*, the assumption of balanced groups, Gaussian distributions, bilateral tests, and $\alpha = 0.05$, for each connection. We found that the estimated sample sizes ranged from 140 to upwards of 600 total subjects, which further suggests that findings from small samples, similar to the seven samples we included when assessed independently, are not expected to be reliable. As noted above, as we used the same sample to estimate the location of effects and

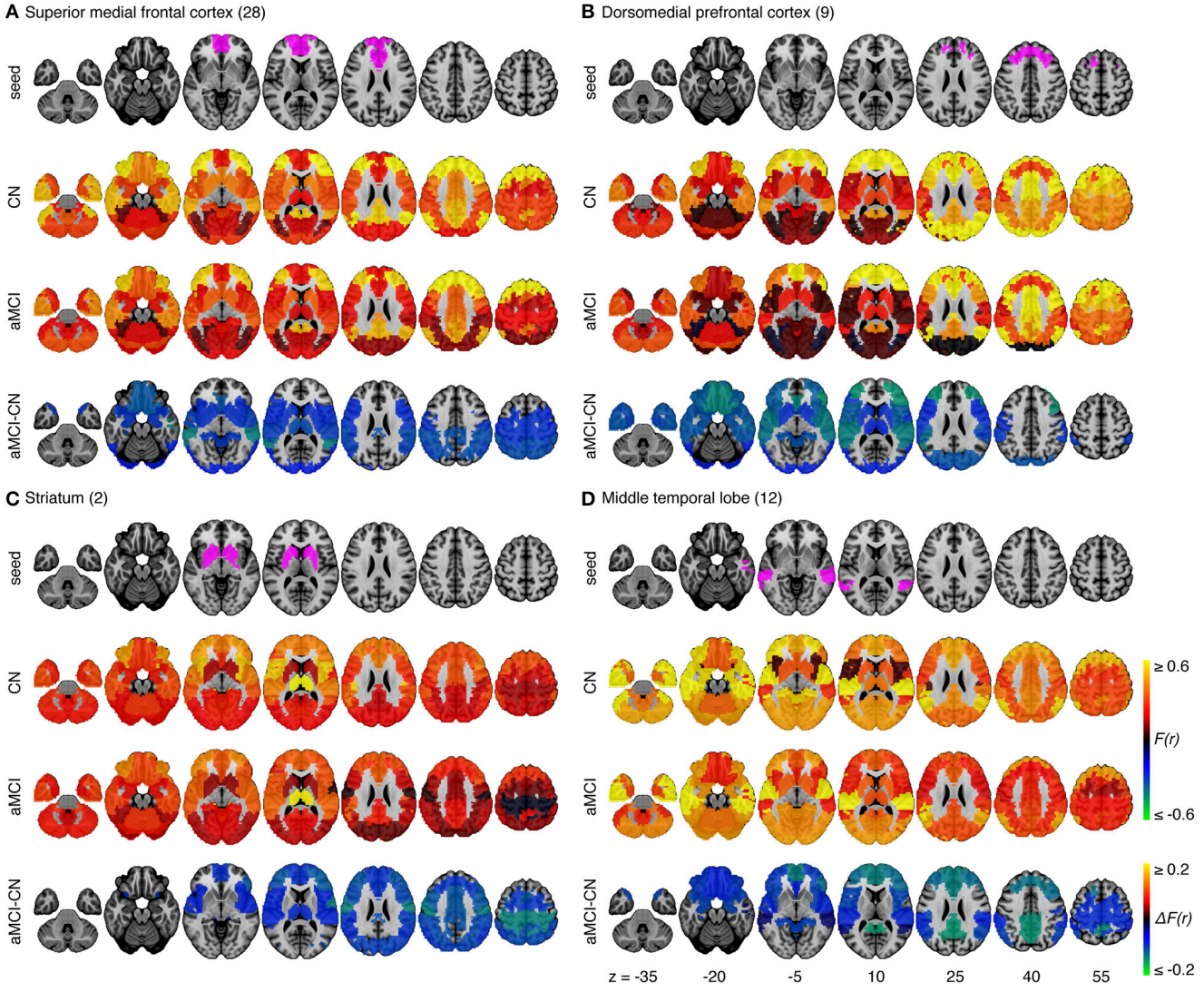

FIGURE 4 | Effect maps for a selection of four seeds that show effects related to aMCI at resolution 33. Effect maps reveal the spatial distribution of the changes in functional connectivity for **(A)** the superior medial frontal cortex, **(B)** the dorsomedial prefrontal cortex, **(C)** striatum, and **(D)** the middle temporal lobe. All connections shown in the maps of difference in average connectivity between aMCI and CN are significant at $q^{FDR} \leq 0.1$. For each panel, the top line maps the spatial location of the seed region in magenta, the second and third lines show the connectivity (Fisher-transformed correlation values, *F(r)*) between the designated seed region and the rest of the brain in CN and aMCI, respectively, and the fourth line shows a difference map between aMCI and CN [difference in Fisher-transformed correlation values, $\Delta F(r)$]. The numbers in parentheses refer to the numerical IDs of the clusters in the 3D parcellation volume, as listed in Supplementary Table 2.

their size, these sample size estimates are possibly optimistic, i.e., deflated compared to a replication on an independent sample. See **Figure 6** and Supplementary Figures 6–9 for Cohen's *d* and sample size estimates for each significant connection that was reported in **Figure 4**.

Effect of Resolution on the GLM

The percentage of discoveries in significant differences between aMCI and CN across the connectomes varied markedly as a function of resolution, as selected by the MSTEPS procedure. Higher resolutions were associated with fewer discoveries, especially beyond resolution 65 (Supplementary Figure 10A). By contrast, the maximal amplitude of differences in average connectivity associated with a particular cluster did not decrease substantially, and sometimes increased, when the resolution increased (Supplementary Figure 10B). The decrease in percentage of discovery thus likely reflected a cost associated with an increased number of multiple comparisons in the FDR procedure, rather than a loss in signal quality. Regarding the clusters that were selected for our seed-based analyses (the superior medial frontal cortex, dorsomedial prefrontal cortex, striatum, and middle temporal lobe), the associated effect maps (without statistical threshold) were highly consistent across different resolutions (Supplementary Figures 11, 12), with the

FIGURE 5 | Comparisons of effects in the superior medial frontal cortex across samples. This figure illustrates functional connectivity changes between aMCI and CN, average connectivity in CN, and average connectivity in aMCI in each site (ADNI2-Achieva, ADNI2-Gemini, ADNI2-Ingenia, ADNI2-Intera, CRIUGMa, CRIUGMb, MNI) independently of other sites and when samples are pooled together (all samples). The number in parentheses refers to the numerical ID of the seed in the 3D parcellation volume, as listed in Supplementary Table 2.

potential exception of very low resolutions where, for example, a relatively small cluster like the anterior cingulate got merged with a large distributed cortical network. This also replicated a prior study on the effect of multiresolution parcellations on GLM analysis (Bellec et al., 2015). Lastly, signal-to-noise ratio did not have a significant impact on the results (Supplementary Figure 13).

DISCUSSION

We report resting-state functional connectivity differences in the superior medial frontal cortex, dorsomedial prefrontal cortex, striatum, and middle temporal lobe between aMCI and CN subjects when multiple studies were combined together. Despite protocol differences, we found that aMCI exhibited reduced connectivity within areas of the DMN and cortico-striatal-thalamic loop compared to CN. Previous studies suggested these altered patterns of functional connectivity in MCI may result from the coevolution of multiple AD-associated biological processes, namely structural degeneration (Pievani et al., 2010; Coupé et al., 2012), neurofibrillary and amyloid pathologies (Small et al., 2006), and cerebrovascular dysfunction (Villeneuve and Jagust, 2015).

The superior medial frontal cortex and middle temporal lobes, both of which are DMN nodes, were among the seed regions with the greatest amount of aMCI-related connectivity changes with other brain areas. Decreased connectivity in aMCI patients was found between these two nodes and other DMN regions, including the posterior cingulate, precuneus, inferior parietal lobes, ventromedial prefrontal cortex, and hippocampus. Our findings support previous studies that used small single-site samples and reported reduced DMN connectivity in MCI and AD patients (Sorg et al., 2007; Bai et al., 2009; Agosta et al., 2012; Koch et al., 2012). Alterations in the DMN may reflect increased amyloid burden in aMCI patients as it has been shown that amyloid plaques impair default mode connectivity (Hedden et al., 2009; Sheline et al., 2010b; Mormino et al., 2011).

We found reduced connectivity within the frontal lobes, notably between ventral and dorsal areas. Decreased functional connectivity between the ventral and dorsal frontal regions could reflect degeneration in gray matter and in white matter tracts connecting these areas. Longitudinal studies have shown greater prefrontal cortex atrophy in MCI over time, as well as in those transitioning to AD, compared to CN (McDonald et al., 2009; Carmichael et al., 2013). Cortico-cortical white matter bundles, e.g., superior longitudinal fasciculus, have also

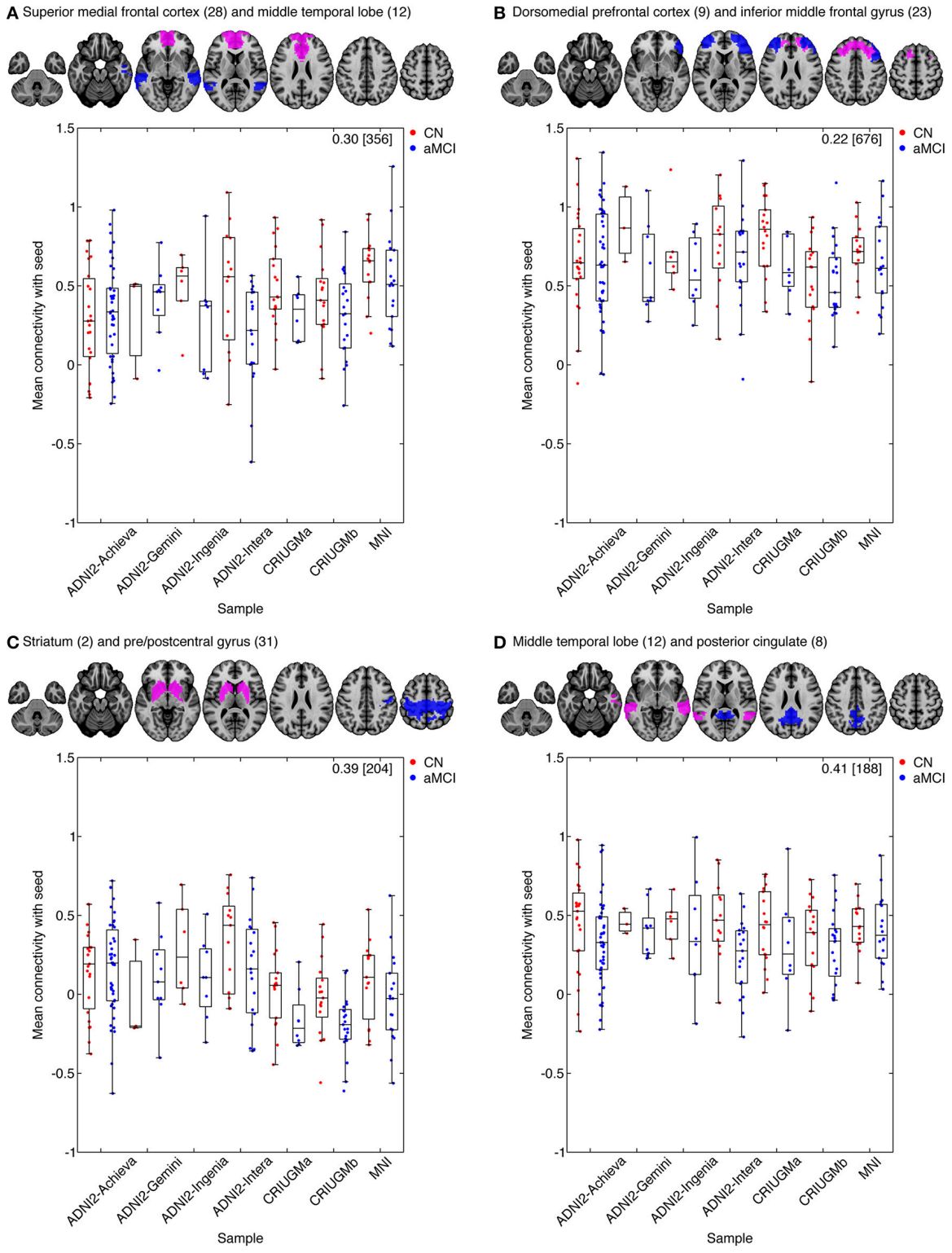

FIGURE 6 | Mean connectivity between (A) the superior medial frontal cortex and middle temporal lobe, (B) the dorsomedial prefrontal cortex and middle frontal gyrus, (C) the striatum and pre/postcentral gyrus, and (D) middle temporal lobe and posterior cingulate in CN and aMCI in the independent samples. Each map displays the seed (pink) and a selected cluster (blue) whose connectivity with the seed significantly differed between CN and aMCI in the pooled analysis. The box-whisker plots display the mean connectivity (Fisher-transformed correlation values) between the seed and the selected parcel, overlaid

(Continued)

FIGURE 6 | Continued
over individual data points, in the CN and MCI groups in the ADNI2-Achieva, ADNI2-Gemini, ADNI2-Ingenia, ADNI2-Intera, CRIUGMa, CRIUGMb, and MNI samples. We also report the Cohen's *d* (a weighted average of the effect sizes per sample) followed by a sample size estimate (for 80% power, balanced groups, bilateral tests, Gaussian distributions, and $\alpha = 0.05$) in square brackets in the top-right corner of each plot. The numbers in parentheses in the titles refer to the numerical IDs of the seeds in the 3D parcellation volume, as listed in Supplementary Table 2. For box-whisker plots for all significant clusters with each of these seeds, see Supplementary Figures 6–9.

been demonstrated to degenerate in patients with MCI and AD (Pievani et al., 2010). Additionally, functional connectivity changes may reflect the regional effect of increased amyloid burden (Sheline et al., 2010b), and PIB-PET work has shown the frontal lobe to be one of the first regions in which amyloid accumulates in autosomal dominant AD mutation carriers (Bateman et al., 2012). Our results may also be due to neurofibrillary pathology as it typically appears in the prefrontal cortex during MCI (Bossers et al., 2010). Lastly, cerebral hypoperfusion in the frontal lobe of MCI (Chao et al., 2009) may have contributed to our results.

We also observed functional disconnection between the temporal and frontal lobes in aMCI. Effects in the temporal lobes were expected given that the temporal lobe is a region known to suffer from significant AD pathology in preclinical phases (Guillozet et al., 2003). Structural connectivity may also explain the functional connectivity changes between the frontal and temporal regions, since degeneration of white matter tracts between these areas, e.g., the uncinate fasciculus, occurs with the progression from MCI to AD and correlates with episodic memory impairment in MCI (Pievani et al., 2010; Rémy et al., 2015). Furthermore, examining the integrity of the arcuate fasciculus, a major language tract that connects the frontal and temporal lobes (Dick and Tremblay, 2012), might reveal a biological basis for language impairments such as word-finding difficulties in MCI and AD, (Nutter-Upham et al., 2008). Brain areas that subserve language function could be important targets to investigate given recent evidence that multilingualism, like other forms of cognitive reserve, may help delay the onset of AD (Chertkow et al., 2010).

Unexpectedly, we also found significant effects in the striatum, which showed reduced connectivity in aMCI with the sensorimotor cortex, frontal and parietal regions, and thalamus. While not initially expected, these findings may reflect earlier observations that regions within the cortico-striatal-thalamic loops are vulnerable to AD pathology. For example, previous work demonstrated the presence of substantial amyloid burden in the striatum in both autosomal dominant and sporadic forms of AD (Braak and Braak, 1990; Villemagne et al., 2009), and the striatum may be the first region in which amyloid deposition occurs in autosomal dominant AD (Klunk et al., 2007; Bateman et al., 2012). Furthermore, significant neurodegeneration is known to occur with AD in the striatum and thalamus (de Jong et al., 2008; Madsen et al., 2010), so our results might reflect the brain's capacity for functional plasticity in response to amyloid or neurodegeneration in these regions. Motor cortex hyperexcitability has also been shown in AD, and this suggests that inhibitory circuits leading to the motor cortex may be affected in the disease (Ferreri et al., 2011). Patients with AD also demonstrate changes in swallowing which have been associated with altered cortical activity (Humbert et al., 2010). Our results may support these observations. Additionally, our findings may represent a biological basis for the cognitive and motor symptoms of MCI (Aggarwal et al., 2006) since the striatum and the rest of the basal ganglia have been implicated in stimulus-response associative learning and memory and motor skill acquisition and execution (Packard and Knowlton, 2002; Doyon et al., 2009). Future research should examine the potential relationship between connectivity in the cortico-striatal-thalamic loops and motor function in aMCI and AD.

Our findings contrasted with previous, smaller single-site studies that have variously reported decreased and increased connectivity. The reports of increased connectivity (Bai et al., 2009; Qi et al., 2010; Gour et al., 2011) may have reflected unique attributes of particular protocols or the choices made with respect to pre-processing steps, for example using global signal regression (Saad et al., 2012). Given that our sample size estimates suggest the use of hundreds of subjects to obtain adequate statistical power, it is not surprising that discrepancies between our results and previous findings generated from smaller, likely underpowered, studies exist. Even when we examined the samples in our study (ADNI2-Achieva, ADNI2-Gemini, ADNI2-Ingenia, ADNI2-Intera, CRIUGMa, CRIUGMb, MNI) independently of each other, we found inconsistent effects among the samples. It is only by combining the studies together in a meta-analysis that we were able to find some common differences in functional connectomes between patients with aMCI and CN. This finding underscores the need for multisite studies with large sample sizes in order to generate reproducible results, as previously suggested in the field of autism research (Haar et al., 2014).

Among our study's limitations is that it was not possible to model each of the 13 ADNI2 sites independently because the sites tended to be small and unbalanced in the numbers of patients and controls. We therefore chose to model each scanner model within ADNI2 separately based on the recommendation of a reviewer. A previous version of the analysis (published as a preprint[9]) had not modeled the different scanner models in ADNI2 and instead treated ADNI2 as a single site. This previous analysis yielded fewer significant findings, but the results were still mostly consistent with what is reported here. Our results suggest that modeling scanner models may have a positive impact on fMRI association studies, but further experiments would be required to confirm that this trend is reproducible. We must also note that the METAL averaging is only representative of the specific samples that were averaged, especially using only

[9]http://dx.doi.org/10.1101/019646.

Philips and Siemens scanners, and it is unclear how our findings may replicate in other studies that would employ a different combination of protocols, say using GE scanners. In particular, our sample size estimates have to be interpreted with caution. They may first be under-estimated, because they were not derived from pre-specified locations, but rather associated with the connections showing the largest effects in our particular sample. These sample sizes were also derived from a meta-analysis combining particular types of studies. We only had 3T scanners from two manufacturers, Siemens and Philips. For the Siemens studies, all were from the same model. For the Philips studies, the scanning protocol was identical at every site, and only the scanner model varied across scanners. Finally, a fairly large number of patients and controls (generally more than 10 subjects per group) was scanned for each variant of the scanning protocol. The sample size estimate may turn out quite differently for a single site study or on the contrary for a study with a very large number of sites and with only a few subjects per site.

Our study is also limited by its cross-sectional nature, which precludes inference that the functional changes we found would necessarily predict progression toward Alzheimer's dementia. Furthermore, aMCI has many underlying causes aside from AD. It is possible that some subjects in our cohort had cognitive impairments due to Lewy Body dementia, for example. However, all samples in the current study had inclusion criteria that enriched for subjects that had aMCI likely due to AD and excluded aMCI subjects with other co-morbidities, such as depression or Parkinson's disease. Also, we did not account for structural atrophy, despite a bias for increased detection in functional differences due to differences in underlying structure (Dukart and Bertolino, 2014). However, aMCI-related gray matter changes likely co-localize to some extent with functional changes, and the aim of our work was to map out functional changes rather than study their interaction with atrophy. We did not account for other variables, such as APOE genotype (Sheline et al., 2010a), amyloid deposition (Sheline et al., 2010b), presence of neurofibrillary tangles (Maruyama et al., 2013), and cerebrovascular mechanisms (Villeneuve and Jagust, 2015). At least some of these could potentially have explained the observed aMCI-related functional connectivity changes as part of an underlying disease mechanism. Large-scale multimodal studies, incorporating genomics, proteomics, and multimodal imaging will be needed to identify the interactions between these and other physiological facets of the pathology. Despite combining several samples together, we still only achieved relatively limited power, given that sample size estimates required at least 140 to over 600 total subjects to consistently identify effects between groups. Lastly, because of the explorative approach used in our study, the resulting estimates of effect sizes may have been inflated and discussion of possible pathological mechanisms for our findings was speculative. However, our discoveries may be used as follow-up targets in future work. Upcoming research should not only attempt to verify our findings by using these regions and their associated connections with hypothesis-driven approaches (e.g., seed-based correlation analyses), but also to extend them to cohorts that include Alzheimer's dementia and other clinical populations (e.g., CN with significant amyloid deposition) and to longitudinal studies that characterize individuals' progression to dementia. Finally, future studies should aim to determine whether our findings are associated with established biomarkers of AD (e.g., amyloid and tau quantification) in order to probe the potential of these functional connections as biomarkers.

Overall, our results supported previous findings of DMN connectivity changes in AD and MCI (Greicius et al., 2004; Sorg et al., 2007), given that three of the identified seeds (superior medial frontal cortex, dorsomedial prefrontal cortex, middle temporal lobe) are part of this network. It is noteworthy, however, that our strongest observed effects reported here were not in the same DMN regions typically described in earlier resting-state studies of MCI and AD, viz, posterior cingulate/precuneus (Sheline et al., 2010b; Zhang et al., 2010). Unexpected changes were also found in the striatum, and this may reflect the advantages of "mining" the whole-brain connectome to search for new biomarkers of mild cognitive impairment and possibly the early progression of the pathophysiologic substrate of Alzheimer's disease. If confirmed, our results could suggest the utility of these regions in resting-state fMRI as a biomarker endpoint in clinical trials.

ACKNOWLEDGMENTS

We thank Sven Joubert, Isabelle Rouleau, and Sophie Benoit for their contribution in collection of the CRIUGMa dataset. We thank Tom Nichols for his comments and suggestions. Collection of the CRIUGMb dataset was supported by the Alzheimer's Society of Canada. Collection of the MNI dataset was supported by Industry Canada/Montreal Neurological Institute Centre of excellence in commercialization and research. This work was supported by CIHR (133359) and NSERC (436141-2013). Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI; National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen Idec Inc.; Bristol-Myers Squibb Company; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd., and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Medpace, Inc.; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Synarc Inc.; and Takeda Pharmaceutical Company. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute

for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fnagi.2015.00242

REFERENCES

Ad-Dab'bagh, Y., Einarson, D., Lyttelton, O., Muehlboeck, J. S., Mok, K., Ivanov, O., et al. (2006). "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* (Florence), 2266.

Aggarwal, N. T., Wilson, R. S., Beck, T. L., Bienias, J. L., and Bennett, D. A. (2006). Motor dysfunction in mild cognitive impairment and the risk of incident Alzheimer disease. *Arch. Neurol.* 63, 1763–1769. doi: 10.1001/archneur.63.12.1763

Agosta, F., Pievani, M., Geroldi, C., Copetti, M., Frisoni, G. B., and Filippi, M. (2012). Resting state fMRI in Alzheimer's disease: beyond the default mode network. *Neurobiol. Aging* 33, 1564–1578. doi: 10.1016/j.neurobiolaging.2011.06.007

Albert, M. S., DeKosky, S. T., Dickson, D., Dubois, B., Feldman, H. H., Fox, N. C., et al. (2011). The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers. Dement.* 7, 270–279. doi: 10.1016/j.jalz.2011.03.008

Bai, F., Watson, D. R., Yu, H., Shi, Y., Yuan, Y., and Zhang, Z. (2009). Abnormal resting-state functional connectivity of posterior cingulate cortex in amnestic type mild cognitive impairment. *Brain Res.* 1302, 167–174. doi: 10.1016/j.brainres.2009.09.028

Bateman, R. J., Xiong, C., Benzinger, T. L. S., Fagan, A., Goate, A., Fox, N. C., et al. (2012). Clinical and biomarker changes in dominantly inherited Alzheimer's disease. *N. Engl. J. Med.* 367, 795–804. doi: 10.1056/NEJMoa1202753

Bellec, P. (2013). "Mining the hierarchy of resting-state brain networks: selection of representative clusters in a multiscale structure," in International Workshop on *Pattern Recognition in Neuroimaging* (Philadelphia, PA), 54–57. doi: 10.1109/prni.2013.23

Bellec, P., Benhajali, Y., Carbonell, F., Dansereau, C., Albouy, G., Pelland, M., et al. (2015). Impact of the resolution of brain parcels on connectome-wide association studies in fMRI. *NeuroImage* 123, 212–228. doi: 10.1016/j.neuroimage.2015.07.071

Bellec, P., Carbonell, F. M., Perlbarg, V., Lepage, C., Lyttelton, O., Fonov, V., et al. (2011). "A neuroimaging analysis kit for Matlab and Octave," in *Proceedings of the 17th International Conference on Functional Mapping of the Human Brain.* (Québec City, QC).

Bellec, P., Lavoie-Courchesne, S., Dickinson, P., Lerch, J. P., Zijdenbos, A. P., and Evans, A. (2012). The pipeline system for Octave and Matlab (PSOM): a lightweight scripting framework and execution engine for scientific workflows. *Front. Neuroinform.* 6:7. doi: 10.3389/fninf.2012.00007

Bellec, P., Perlbarg, V., Jbabdi, S., Pélégrini-Issac, M., Anton, J.-L., Doyon, J., et al. (2006). Identification of large-scale networks in the brain using fMRI. *Neuroimage* 29, 1231–1243. doi: 10.1016/j.neuroimage.2005.08.044

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* 57, 289–300. doi: 10.2307/2346101

Bossers, K., Wirz, K. T. S., Meerhoff, G. F., Essing, A. H. W., van Dongen, J. W., Houba, P., et al. (2010). Concerted changes in transcripts in the prefrontal cortex precede neuropathology in Alzheimer's disease. *Brain* 133, 3699–3723. doi: 10.1093/brain/awq258

Braak, H., and Braak, E. (1990). Alzheimer's disease: striatal amyloid deposits and neurofibrillary changes. *J. Neuropathol. Exp. Neurol.* 49, 215. doi: 10.1097/00005072-199005000-00003

Carmichael, O., McLaren, D. G., Tommet, D., Mungas, D., Jones, R. N., for the Alzheimer's Disease Neuroimaging Initiative (2013). Coevolution of brain structures in amnestic mild cognitive impairment. *NeuroImage* 66, 449–456. doi: 10.1016/j.neuroimage.2012.10.029

Chao, L. L., Pa, J., Duarte, A., Schuff, N., Weiner, M. W., Kramer, J. H., et al. (2009). Patterns of cerebral hypoperfusion in amnestic and dysexecutive MCI. Alzheimer Dis. Assoc. Disord. 23, 245–252. doi: 10.1097/WAD.0b013e318199ff46

Chertkow, H., Whitehead, V., Phillips, N., Wolfson, C., Atherton, J., and Bergman, H. (2010). Multilingualism (but not always bilingualism) delays the onset of alzheimer disease: evidence from a bilingual community. Alzheimer Dis. Assoc. Disord. 24, 118–125. doi: 10.1097/WAD.0b013e3181ca1221

Collins, D. L., and Evans, A. C. (1997). Animal: validation and applications of nonlinear registration-based segmentation. *Int. J. Patt. Recogn. Artif. Intell.* 11, 1271–1294. doi: 10.1142/S0218001497000597

Coupé, P., Eskildsen, S. F., Manjón, J. V., Fonov, V. S., Collins, D. L., for the Alzheimer's Disease Neuroimaging Initiative (2012). Simultaneous segmentation and grading of anatomical structures for patient"s classification: application to Alzheimer"s disease. *NeuroImage* 59, 3736–3747. doi: 10.1016/j.neuroimage.2011.10.080

de Jong, L. W., van der Hiele, K., Veer, I. M., Houwing, J. J., Westendorp, R. G. J., Bollen, E. L. E. M., et al. (2008). Strongly reduced volumes of putamen and thalamus in Alzheimer's disease: an MRI study. *Brain* 131, 3277–3285. doi: 10.1093/brain/awn278

Dick, A. S., and Tremblay, P. (2012). Beyond the arcuate fasciculus: consensus and controversy in the connectional anatomy of language. *Brain* 135, 3529–3550. doi: 10.1093/brain/aws222

Doyon, J., Bellec, P., Amsel, R., Penhune, V., Monchi, O., Carrier, J., et al. (2009). Contributions of the basal ganglia and functionally related brain structures to motor learning. *Behav. Brain Res.* 199, 61–75. doi: 10.1016/j.bbr.2008.11.012

Dukart, J., and Bertolino, A. (2014). When structure affects function – the need for partial volume effect correction in functional and resting state magnetic resonance imaging studies. *PLoS ONE* 9:e114227. doi: 10.1371/journal.pone.0114227

Edward, V., Windischberger, C., Cunnington, R., Erdler, M., Lanzenberger, R., Mayer, D., et al. (2000). Quantification of fMRI artifact reduction by a novel plaster cast head holder. *Hum. Brain Mapp.* 11, 207–213. doi: 10.1002/1097-0193(200011)11:3<207::AID-HBM60>3.0.CO;2-J

Elliott, M. R., Bowtell, R. W., and Morris, P. G. (1999). The effect of scanner sound in visual, motor, and auditory functional MRI. *Magn. Reson. Med.* 41, 1230–1235.

Ferreri, F., Pasqualetti, P., Määttä, S., Ponzo, D., Guerra, A., Bressi, F., et al. (2011). Motor cortex excitability in Alzheimer's disease: a transcranial magnetic stimulation follow-up study. *Neurosci. Lett.* 492, 94–98. doi: 10.1016/j.neulet.2011.01.064

Fonov, V., Evans, A., Botteron, K., Almli, C. R., McKinstry, R. C., and Collins, D. L. (2011). Unbiased average age-appropriate atlases for pediatric studies. *NeuroImage* 54, 313–327. doi: 10.1016/j.neuroimage.2010.07.033

Friedman, L., and Glover, G. H. (2006). Report on a multicenter fMRI quality assurance protocol. *J. Magn. Reson. Imaging* 23, 827–839. doi: 10.1002/jmri.20583

Friedman, L., Glover, G. H., and Fbirn Consortium (2006). Reducing interscanner variability of activation in a multicenter fMRI study: controlling for signal-to-fluctuation-noise-ratio (SFNR) differences. *NeuroImage* 33, 471–481. doi: 10.1016/j.neuroimage.2006.07.012

Giove, F., Gili, T., Iacovella, V., Macaluso, E., and Maraviglia, B. (2009). Images-based suppression of unwanted global signals in resting-state functional connectivity studies. *Magn. Reson. Imaging* 27, 1058–1064. doi: 10.1016/j.mri.2009.06.004

Gour, N., Ranjeva, J.-P., Ceccaldi, M., Confort-Gouny, S., Barbeau, E., Soulier, E., et al. (2011). Basal functional connectivity within the anterior temporal network is associated with performance on declarative memory tasks. *NeuroImage* 58, 687–697. doi: 10.1016/j.neuroimage.2011.05.090

Greicius, M. D., Srivastava, G., Reiss, A. L., and Menon, V. (2004). Default-mode network activity distinguishes Alzheimer's disease from healthy aging:

evidence from functional MRI. *Proc. Natl. Acad. Sci. U.S.A.* 101, 4637–4642. doi: 10.1073/pnas.0308627101

Guillozet, A. L., Weintraub, S., Mash, D. C., and Mesulam, M. M. (2003). Neurofibrillary tangles, amyloid, and memory in aging and mild cognitive impairment. *Arch. Neurol.* 60, 729–736. doi: 10.1001/archneur.60.5.729

Haar, S., Berman, S., Behrmann, M., and Dinstein, I. (2014). Anatomical Abnormalities in Autism? *Cereb. Cortex*. doi: 10.1093/cercor/bhu242. [Epub ahead of print].

Hedden, T., van Dijk, K. R. A., Becker, J. A., Sperling, R. A., Johnson, K. A., and Buckner, R. L. (2009). Disruption of functional connectivity in clinically normal older adults harboring amyloid burden. *J. Neurosci.* 29, 12686–12694. doi: 10.1523/JNEUROSCI.3189-09.2009

Humbert, I. A., McLaren, D. G., Kosmatka, K., Fitzgerald, M., Johnson, S., Porcaro, E., et al. (2010). Early deficits in cortical control of swallowing in Alzheimer's disease. *J. Alzheimers Dis.* 19, 1185–1197. doi: 10.3233/JAD-2010-1316

Kelly, C., Biswal, B. B., Craddock, R. C., Castellanos, F. X., and Milham, M. P. (2012). Characterizing variation in the functional connectome: promise and pitfalls. *Trends Cogn. Sci.* 16, 181–188. doi: 10.1016/j.tics.2012.02.001

Klunk, W. E., Price, J. C., Mathis, C. A., Tsopelas, N. D., Lopresti, B. J., Ziolko, S. K., et al. (2007). Amyloid deposition begins in the striatum of presenilin-1 mutation carriers from two unrelated pedigrees. *J. Neurosci.* 27, 6174–6184. doi: 10.1523/JNEUROSCI.0730-07.2007

Koch, W., Teipel, S., Mueller, S., Benninghoff, J., Wagner, M., Bokde, A. L. W., et al. (2012). Diagnostic power of default mode network resting state fMRI in the detection of Alzheimer's disease. *Neurobiol. Aging* 33, 466–478. doi: 10.1016/j.neurobiolaging.2010.04.013

Liang, P., Wang, Z., Yang, Y., and Li, K. (2012). Three subsystems of the inferior parietal cortex are differently affected in mild cognitive impairment. *J. Alzheimers Dis.* 30, 475–487. doi: 10.3233/JAD-2012-111721

Lund, T. E., Madsen, K. H., Sidaros, K., Luo, W.-L., and Nichols, T. E. (2006). Non-white noise in fMRI: does modelling have an impact? *Neuroimage* 29, 54–66. doi: 10.1016/j.neuroimage.2005.07.005

Madsen, S. K., Ho, A. J., Hua, X., Saharan, P. S., Toga, A. W., Jack, C. R., et al. (2010). 3D maps localize caudate nucleus atrophy in 400 Alzheimer's disease, mild cognitive impairment, and healthy elderly subjects. *Neurobiol. Aging* 31, 1312–1325. doi: 10.1016/j.neurobiolaging.2010.05.002

Maruyama, M., Shimada, H., Suhara, T., Shinotoh, H., Ji, B., Maeda, J., et al. (2013). Imaging of tau pathology in a tauopathy mouse model and in alzheimer patients compared to normal controls. *Neuron* 79, 1094–1108. doi: 10.1016/j.neuron.2013.07.037

McDonald, C. R., McEvoy, L. K., Gharapetian, L., Fennema-Notestine, C., Hagler, D. J., Holland, D., et al. (2009). Regional rates of neocortical atrophy from normal aging to early Alzheimer disease. *Neurology* 73, 457–465. doi: 10.1212/WNL.0b013e3181b16431

Mormino, E. C., Smiljic, A., Hayenga, A. O., Onami, S. H., Greicius, M. D., Rabinovici, G. D., et al. (2011). Relationships between β-amyloid and functional connectivity in different components of the default mode network in aging. *Cereb. Cortex* 21, 2399–2407. doi: 10.1093/cercor/bhr025

Nutter-Upham, K. E., Saykin, A. J., Rabin, L. A., Roth, R. M., Wishart, H. A., Pare, N., et al. (2008). Verbal fluency performance in amnestic MCI and older adults with cognitive complaints. *Arch. Clin. Neuropsychol.* 23, 229–241. doi: 10.1016/j.acn.2008.01.005

Packard, M. G., and Knowlton, B. J. (2002). Learning and memory functions of the Basal Ganglia. *Annu. Rev. Neurosci.* 25, 563–593. doi: 10.1146/annurev.neuro.25.112701.142937

Petersen, R. C. (2004). Mild cognitive impairment as a diagnostic entity. *J. Intern. Med.* 256, 183–194. doi: 10.1111/j.1365-2796.2004.01388.x

Petersen, R. C., Doody, R., Kurz, A., Mohs, R. C., Morris, J. C., Rabins, P. V., et al. (2001). Current concepts in mild cognitive impairment. *Arch. Neurol.* 58, 1985–1992. doi: 10.1001/archneur.58.12.1985

Pievani, M., Agosta, F., Pagani, E., Canu, E., Sala, S., Absinta, M., et al. (2010). Assessment of white matter tract damage in mild cognitive impairment and Alzheimer's disease. *Hum. Brain Mapp.* 31, 1862–1875. doi: 10.1002/hbm.20978

Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L., and Petersen, S. E. (2012). Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *Neuroimage* 59, 2142–2154. doi: 10.1016/j.neuroimage.2011.10.018

Qi, Z., Wu, X., Wang, Z., Zhang, N., Dong, H., Yao, L., et al. (2010). Impairment and compensation coexist in amnestic MCI default mode network. *Neuroimage* 50, 48–55. doi: 10.1016/j.neuroimage.2009.12.025

Rémy, F., Vayssière, N., Saint-Aubert, L., Barbeau, E., and Pariente, J. (2015). White matter disruption at the prodromal stage of Alzheimer's disease: relationships with hippocampal atrophy and episodic memory performance. *Neuroimage Clin.* 7, 482–492. doi: 10.1016/j.nicl.2015.01.014

Saad, Z. S., Gotts, S. J., Murphy, K., Chen, G., Jo, H. J., Martin, A., et al. (2012). Trouble at rest: how correlation patterns and group differences become distorted after global signal regression. *Brain Connect.* 2, 25–32. doi: 10.1089/brain.2012.0080

Shehzad, Z., Kelly, C., Reiss, P. T., Cameron Craddock, R., Emerson, J. W., McMahon, K., et al. (2014). A multivariate distance-based analytic framework for connectome-wide association studies. *NeuroImage* 93(Pt 1), 74–94. doi: 10.1016/j.neuroimage.2014.02.024

Sheline, Y. I., Morris, J. C., Snyder, A. Z., Price, J. L., Yan, Z., D'Angelo, G., et al. (2010a). APOE4 allele disrupts resting state fMRI connectivity in the absence of amyloid plaques or decreased CSF A 42. *J. Neurosci.* 30, 17035–17040. doi: 10.1523/JNEUROSCI.3987-10.2010

Sheline, Y. I., Raichle, M. E., Snyder, A. Z., Morris, J. C., Head, D., Wang, S., et al. (2010b). Amyloid plaques disrupt resting state default mode network connectivity in cognitively normal elderly. *Biol. Psychiatry* 67, 584–587. doi: 10.1016/j.biopsych.2009.08.024

Small, G. W., Kepe, V., Ercoli, L. M., Siddarth, P., Bookheimer, S. Y., Miller, K. J., et al. (2006). PET of brain amyloid and tau in mild cognitive impairment. *N. Engl. J. Med.* 355, 2652–2663. doi: 10.1056/NEJMoa054625

Sorg, C., Riedl, V., Mühlau, M., Calhoun, V. D., Eichele, T., Läer, L., et al. (2007). Selective changes of resting-state networks in individuals at risk for Alzheimer's disease. *Proc. Natl. Acad. Sci. U.S.A.* 104, 18760–18765. doi: 10.1073/pnas.0708803104

van Dijk, K. R. A., Hedden, T., Venkataraman, A., Evans, K. C., Lazar, S. W., and Buckner, R. L. (2010). Intrinsic functional connectivity as a tool for human connectomics: theory, properties, and optimization. *J. Neurophysiol.* 103, 297–321. doi: 10.1152/jn.00783.2009

Vanhoutte, G., Verhoye, M., and van der Linden, A. (2006). Changing body temperature affects the T2* signal in the rat brain and reveals hypothalamic activity. *Magn. Reson. Med.* 55, 1006–1012. doi: 10.1002/mrm.20861

Villemagne, V. L., Ataka, S., Mizuno, T., Brooks, W. S., Wada, Y., Kondo, M., et al. (2009). High striatal amyloid beta-peptide deposition across different autosomal Alzheimer disease mutation types. *Arch. Neurol.* 66, 1537–1544. doi: 10.1001/archneurol.2009.285

Villeneuve, S., and Jagust, W. J. (2015). Imaging vascular disease and amyloid in the aging brain: implications for treatment. *J. Prev. Alzheimers Dis.* 2, 64–70. doi: 10.14283/jpad.2015.47

Willer, C. J., Li, Y., and Abecasis, G. R. (2010). METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 26, 2190–2191. doi: 10.1093/bioinformatics/btq340

Wu, L., Soder, R. B., Schoemaker, D., Carbonnell, F., Sziklas, V., Rowley, J., et al. (2014). Resting state executive control network adaptations in amnestic mild cognitive impairment. *J. Alzheimers Dis.* 40, 993–1004. doi: 10.3233/JAD-131574

Yan, C.-G., Liu, D., He, Y., Zou, Q., Zhu, C., Zuo, X., et al. (2009). Spontaneous Brain activity in the default mode network is sensitive to different resting-state conditions with limited cognitive load. *PLoS ONE* 4:e5743. doi: 10.1371/journal.pone.0005743

Zhang, H.-Y., Wang, S.-J., Liu, B., Ma, Z.-L., Yang, M., Zhang, Z., et al. (2010). Resting brain connectivity: changes during the progress of Alzheimer disease. *Radiology* 256, 598–606. doi: 10.1148/radiol.10091701

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Appendix II

Second Appendix

Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring 8 (2017) 73-85

Neuroimaging

Resting-state network dysfunction in Alzheimer's disease: A systematic review and meta-analysis

AmanPreet Badhwar[a,b,*], Angela Tam[a,c,d], Christian Dansereau[a,b], Pierre Orban[a,b,d], Felix Hoffstaedter[e,f,g], Pierre Bellec[a,b,**]

[a]Centre de Recherche, Institut Universitaire de Gériatrie de Montréal, Montreal, Quebec, Canada
[b]Université de Montréal, Montreal, Quebec, Canada
[c]McGill University, Montreal, Quebec, Canada
[d]Douglas Mental Health University Institute Research Centre, Montreal, Quebec, Canada
[e]Institute of Neuroscience and Medicine (INM-1, INM-7), Research Centre Jülich, Jülich, Germany
[f]Institute of Clinical Neuroscience and Medical Psychology, Heinrich Heine University Düsseldorf, Düsseldorf, Germany
[g]Institute of Systems Neuroscience, Heinrich Heine University Düsseldorf, Düsseldorf, Germany

Abstract

Introduction: We performed a systematic review and meta-analysis of the Alzheimer's disease (AD) literature to examine consistency of functional connectivity alterations in AD dementia and mild cognitive impairment, using resting-state functional magnetic resonance imaging.

Methods: Studies were screened using a standardized procedure. Multiresolution statistics were performed to assess the spatial consistency of findings across studies.

Results: Thirty-four studies were included (1363 participants, average 40 per study). Consistent alterations in connectivity were found in the default mode, salience, and limbic networks in patients with AD dementia, mild cognitive impairment, or in both groups. We also identified a strong tendency in the literature toward specific examination of the default mode network.

Discussion: Convergent evidence across the literature supports the use of resting-state connectivity as a biomarker of AD. The locations of consistent alterations suggest that highly connected hub regions in the brain might be an early target of AD.

Keywords: Resting-state fMRI; Functional connectivity; Alzheimer's disease; Mild cognitive impairment; Meta-analysis

1. Introduction

Alzheimer's disease (AD) exists on a continuum comprising a lengthy preclinical stage, a middle stage of mild cognitive impairment (MCI), and a final stage of dementia [1]. Symptoms usually start around the age of 65 years, except in rare patients with early onset (33–60 years) autosomal dominant AD (ADAD) [2,3]. Drugs currently available for AD provide limited short-term treatment of AD symptoms [4]. Trials of disease-modifying therapies for AD dementia patients have been unsuccessful, likely because intervention at this stage is too late to affect the neurodegenerative process. The focus now is on therapeutic intervention at the MCI and/or preclinical disease stages, with delay of dementia onset constituting a major clinical end point for clinical trials [1]. This approach depends on the identification of biomarkers that can aid early AD diagnosis [1,5]. Currently, validated AD biomarkers are (1) low cerebrospinal fluid (CSF) amyloid-β 42 levels and/or high

*Corresponding author. Tel.: +1-514-340-3540x3367; Fax: +1-514-340-2802.
**Corresponding author. Tel.: +1-514-340-3540x4782; Fax: +1-514-340-2802.
E-mail address: amanpreet.badhwar@criugm.qc.ca (A.B.), pierre.bellec@criugm.qc.ca (P.B.)

http://dx.doi.org/10.1016/j.dadm.2017.03.007

amyloid tracer retention on positron emission tomography (PET), indicating brain amyloidosis; (2) high CSF tau levels, indicating neuronal injury; (3) temporoparietal pattern of reduced 18F-fluorodeoxyglucose uptake on PET, indicating brain hypometabolism, and (4) patterns of brain atrophy on structural magnetic resonance imaging (MRI), indicating neurodegeneration [1,6].

Connectivity in resting-state functional magnetic resonance imaging (rsfMRI) is an emerging AD biomarker that holds promise for early diagnosis [1,5,7]. RsfMRI indirectly measures neural processing in the brain using blood oxygenation and can be used to identify spatially distributed networks [8]. The National Institute on Aging–Alzheimer's Association lists rsfMRI functional connectivity as a potential biomarker of neuronal injury, at an early stage of validation [6]. The existing literature is indeed mostly composed of proof-of-concept cross-sectional comparisons of cognitively healthy elderly individuals with patients suffering from mild (MCI) or severe (dementia) AD symptoms.

To date, multiple studies have reported intrinsic connectivity network (ICN) disturbances in patients with AD dementia and MCI, presymptomatic ADAD mutation carriers, and cognitively normal individuals carrying the at-risk APOEε4 allele and/or showing evidence of amyloidosis [9–12]. Despite such promising findings, the overall effect of AD on ICNs remains poorly characterized because of several inconsistencies in the literature, such as different acquisition protocols, processing methods, and/or exclusion/inclusion criteria [13]. Our aim was to perform a systematic review and meta-analysis to examine the consistency of intrinsic connectivity alterations in MCI and late-onset AD (LOAD) dementia across the literature. We also reviewed the burgeoning literature on connectivity abnormalities in ADAD and the at-risk APOEε4 genotype.

2. Methods

2.1. Literature search

We conducted a systematic review of PubMed articles up to December 3, 2015 in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines [14]. Search terms and combinations used are provided in Supplementary Table 1. Results were filtered for duplicates within each of the two main search categories, that is, AD dementia or MCI patients (Fig. 1). Unique search results underwent further screening as described subsequently.

2.2. Study selection

Search results were subjected to two successive screenings with increasingly stringent criteria. The initial screen was performed on article abstracts. An article was included if the abstract indicated that it was a peer-reviewed original research article written in English and used rsfMRI to study LOAD and/or MCI in humans. Reviews, letters, case reports, and studies with subjects in whom MCI was associated with other diseases were omitted. Following the initial screening, we applied the following inclusion criteria: (1) used seed-based or independent component analysis rsfMRI methods; (2) investigated functional connectivity between patients (AD dementia or MCI) and age-matched healthy controls (HC); and (3) reported peak coordinates of significant statistical differences in average connectivity between groups and the direction of difference.

2.3. Data extraction

One reviewer (A.B.) conducted the searches and screened for duplicates. Two reviewers (A.B. and A.T.) independently screened all unique search results for potential inclusion in the meta-analysis. Only articles passing both reviewers' approval were considered for final inclusion. For each "included" article, coordinate data of significant between-group comparisons, such as AD versus HC, were transcribed by one reviewer and checked by two others (second reviewer [A.T.] and F.H.).

2.4. Meta-analysis

We performed complementary network- and voxel-based quantitative meta-analyses on six main group comparisons: pooled group with AD dementia and MCI patients termed ADMCI $<$ HC, ADMCI $>$ HC, MCI $<$ HC, MCI $>$ HC, AD $<$ HC, and AD $>$ HC. Although the voxel-based meta-analysis has finer spatial resolution for findings with high anatomic consistency, we assumed the network-based approach would have better sensitivity for detecting consistent involvement of anatomically distributed networks. Coordinates from articles using the same cohort were pooled under the PubMed unique identifier or PMID of the earliest publication and treated as results from a single study to avoid counting the cohort multiple times. Henceforth, an individual article will be referred to as a "study" and a group comparison yielding network and/or localization information (e.g., ADMCI $<$ HC) as a "contrast."

2.4.1. Network-based statistics

We performed network-based statistics on seed coordinates (seed statistics) to assess whether seed regions were preferentially selected from within certain networks in the literature. We also performed network-based statistics on coordinate data of significant contrasts (contrast statistics) to assess the consistency of network-level findings in the AD literature. In particular, we performed three types of contrast statistics: (1) all coordinates irrespective of seed network; and given the focus on the default mode network (DMN) in the literature, (2) coordinates associated with seeds inside the DMN only; and (3) coordinates associated with seeds outside the DMN, that is, non-DMN seeds. All analyses were conducted using a multiresolution atlas of group-level

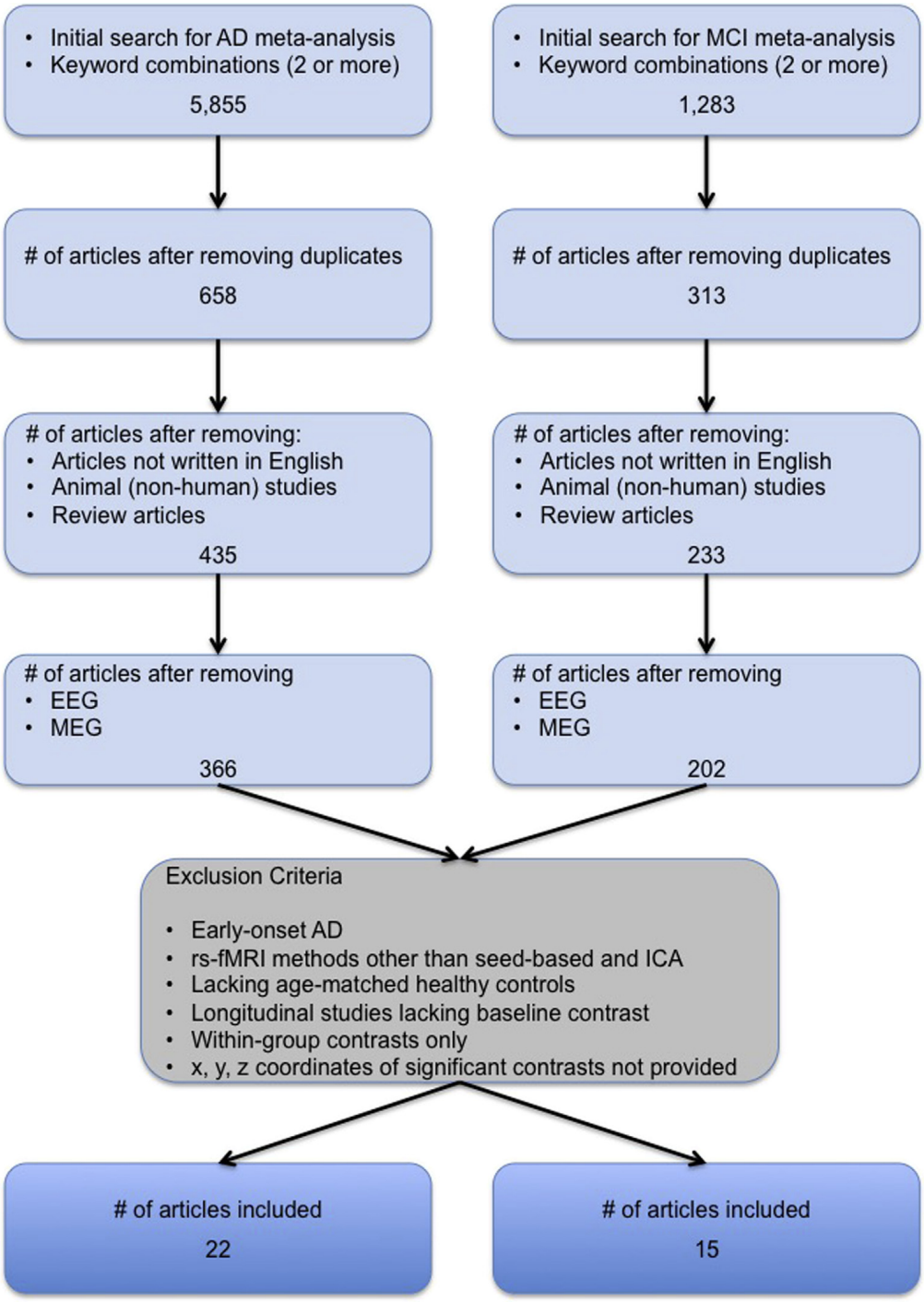

Fig. 1. Flowchart of the study selection process. Selection process for AD and MCI studies included in the meta-analyses. Studies using rsfMRI methods dissimilar to seed-based and ICA methods, such as degree centrality or graph theory, amplitude of low-frequency fluctuations, and regional homogeneity were not included. Abbreviations: AD, Alzheimer's disease; EEG, electroencephalogram; ICA, independent component analysis; MCI, mild cognitive impairment; MEG, magnetoencephalography; rsfMRI, resting-state functional magnetic resonance imaging.

functional brain parcellations derived from an independent rsfMRI data set, the Bootstrap Analysis of Stable Clusters–Cambridge atlas (https://dx.doi.org/10.6084/m9.figshare.1285615.v1) [15]. This atlas consists of nine functional parcellations capturing successively finer levels of spatial detail, of which we used parcellations at two resolutions: the first comprised seven commonly used large-scale networks (R7 atlas) and the second containing 36 networks (R36 atlas). We used R7 and R36 atlases for contrast statistics and only the R7 atlas for seed statistics. Because seeds were assigned indirectly for studies where coordinates were not provided, indirect assignment could not be performed with sufficient precision to use the R36 atlas. Assignment of seeds to one of the R7 networks was based on published coordinates,

when available. When only anatomic labels were provided for seed regions, network assignment was based on (1) the center of gravity in MNI space or (2) visual approximation if no further information was available. For independent component analysis-based studies, network assignment was based on (1) network coordinates when provided or (2) visual assignment to one or more of the seven networks based on the degree of spatial overlap.

We tested the spatial consistency of both seed and peak locations using the following approach. For each study, we computed the number of coordinates falling within each network, after conversion of Talairach space coordinates into MNI space using the Lancaster transform [16], when necessary. Coordinates falling outside of the gray matter mask (ICBM152) were assigned to the closest network. To remain unbiased to the number of coordinates reported per study, we computed the ratio of coordinates falling within each network to the total number of coordinates reported per study. This ratio was then averaged across studies. The significance of findings was assessed using Monte Carlo permutation tests. Using the total number of coordinates per study, we generated a random assignment of coordinates to networks, taking into consideration the volume of each network. Coordinate counts per network were normalized as described previously, followed by an averaging across studies. This Monte Carlo sampling process was repeated 10,000 times. Thereafter, we compared the distribution of the average frequency obtained from the random sampling with the frequency obtained from the meta-analysis, resulting in P value estimates [17]. Multiple comparisons across networks were accounted for using a false discovery rate (FDR) procedure (qFDR < 0.05) [18]. The P values less than .05 that did not survive multiple comparisons were deemed as "trends."

2.4.2. Voxel-based statistics

Voxel-level statistical analysis was performed using activation likelihood estimation (ALE), a widely used algorithm for coordinate-based meta-analysis of neuroimaging studies. ALE aims at delineating brain regions with above-chance convergence of reported coordinates across experiments [19]. Coordinates falling outside the gray matter mask were removed from the analysis. We used the in-house ALE algorithm implementation in MATLAB version 8.3.0.532, which treats each of the coordinates in a given experiment as a three-dimensional gaussian probability distribution centered at the given coordinate. The probability distributions acknowledge the spatial uncertainty associated with each experiment. For any given study, the width of the spatial uncertainty of its coordinates is determined based on empirical data on the between-subject and between-template variances representing the main components of this uncertainty [19]. Then, the probability distributions of all coordinates per included study are combined for each voxel, generating a modeled activation (MA) map. To limit the effect of multiple coordinates very close to one another within a given study, we used the "nonadditive" approach, which calculates MA maps by taking the maximum probability across overlapping gaussians [19]. ALE scores were computed on a voxel-by-voxel basis by taking the union across these MA maps. To distinguish between "true" and random convergence between studies (i.e., noise), ALE scores were compared with a null distribution reflecting a random spatial association between experiments (10,000 permutations). Nonparametric P values were assessed at a familywise error-corrected threshold of $P < .05$ on a cluster level (cluster-forming threshold: $P < .001$ at voxel level) and transformed into t scores for display purposes. Only contrasts including more than 18 experiments were considered, as recommended in a recent large-scale simulation study [20].

3. Results

3.1. Search results

The results of the initial search, along with studies systematically excluded from inclusion in our rsfMRI meta-analyses are presented in Fig. 1. Thirty-four studies totaling 1363 subjects (post pooling of identical cohorts) met our inclusion criteria and were included in the meta-analysis. The total included 352 MCI, 378 AD dementia (specifically LOAD), and 633 HC. Diagnostic criteria used per study for MCI and AD dementia are provided in the Supplementary Material (Supplementary Table 2 and Section 2). The bulk (54%) of the studies had 20 or less subjects per group. Twenty studies (66.7%) investigated rsfMRI connectivity measures with other domains, cognition being most frequent ($n = 11/22$ AD studies, $n = 9/15$ MCI studies), and few with levels of amyloid burden using Pittsburgh compound B ($n = 3$), brain atrophy ($n = 3$), and structural connectivity ($n = 1$). Alterations in functional connectivity were often ($n = 5/9$ studies) reported to be significantly correlated with episodic verbal learning and memory in MCI cohorts. Table 1 provides additional characteristics of the included rsfMRI studies, including scanner make, model, and strength, and seed region and/or ICN investigated. A summary of commonly used preprocessing steps utilized by the studies present in our meta-analysis are provided in Supplementary Table 3.

3.2. Network-based meta-analysis

3.2.1. Seed statistics

Using network-level statistics, we demonstrated that a disproportionately large number of studies specifically targeted the DMN (Fig. 2) irrespective of the population (ADMCI, MCI, or AD dementia) being studied.

3.2.2. Contrast statistics

We first examined R7 network-level statistics and all seeds combined. Aberrant functional brain connectivity was observed in ADMCI, MCI, and AD, relative to HC (Fig. 3). In the ADMCI cohort, we found both significant

Table 1
Characteristics of rsfMRI studies included in the meta-analysis

Study	*N*	AD					HC					AD		Scanner	Method	Seed region/ICN investigated
		n	M	F	Age	SD	*n*	M	F	Age	SD	<HC	>HC			
Wang et al. [63]	28	14	7	7	70.2	6.3	14	7	7	69.6	5.5	x		1.5 T S	SB	PCC
Zhang et al. [64][a]	32	16	6	10	71.6	5.1	16	7	9	71.3	4.9	x	x	1.5 T P	SB	PCC
Zhang et al. [65][a]	55	39	18	21	73.4		16	7	9	71.3	4.9	x	x	1.5 T P	SB	PCC
Sheline et al. [66]	83	35					48					x	x	3.0 T S	SB	Precuneus
Zhou et al. [52]	24	12	5	7	63.3	7.7	12	5	7	62.0		x	x	1.5/3.0/4.0 T S/GE/B	*SB* and ICA	AG (l), pregenual ACC (r)
Gili et al. [67]*	21	11	7	4	71.9	7.9	10	7	3	64.1	10.5	x		3.0 T S	*SB* and ICA	PCC, mPFC
Wu et al. [68][b]	31	15	6	9	64.0	8.3	16	7	9	65.0	9.2	x		3.0 T S	ICA	DMN
Li et al. [69][b]	31	15	6	9	64.0	8.3	16	7	9	65.0	9.2	x		3.0 T S	ICA	ATN (d, v)
Damoiseaux et al. [70]	39	21	9	12	64.2	8.7	18	12	6	62.7	10.3	x	x	3.0 T GE	ICA	DMN (a, p, v), SMN
Binnewijzend et al. [71]†	82	39	23	16	67.0	8.0	43	23	20	69.0	7.0	x		1.5 T S	ICA	DMN, working memory (l, r), visuospatial attention (d), spatial attention (v), SMN, auditory language, prVIS, sVIS, basal ganglia cerebellum
Kenny et al. [72]	32	16			77.3	8.9	16			76.3	8.3		x	3.0 T P	SB	Hippocampus (l, r), PCC, precuneus, prVIS
Zhu et al. [73]†	22	10	7	3	72.9	7.9	12	5	7	73.8	6.5	x		3.0 T GE	SB	ICC (l, r)
Balthazar et al. [74]	37	20			73.9	8.2	17			72.3	6.4	x	x	3.0 T P	ICA	DMN (d, v), SN (a, p)
Yao et al. [75][c,†]	62	35	12	23	72.4	8.5	27	16	11	69.2	6.5	x		3.0 T GE	SB	Amygdala (l, r)
Zhou et al. [76][c,†]	62	35	12	23	72.4	8.5	27	16	11	69.2	6.5	x	x	3.0 T GE	SB	T
Zhang et al. [77][c,†]	62	35	12	23	72.4	8.5	27	16	11	69.2	6.5	x		3.0 T GE	SB	MrD (l, r)
Gour et al. [29]	28	14	6	8	75.1	2.9	14	4	10	72.8	3.0	x	x	3.0 T S	SB	PCC, perirhinal cortex (l, r), dlPFC (l, r)
Weiler et al. [78]	48	22	6	16	73.4	5.7	26	6	20	70.0	6.6	x	x	3.0 T P	SB	PCC, Wernicke's (l); Broca's (l), dlPFC (l, r), saVC
Balachandar et al. [79]	30	15	9	6	67.3	6.6	15	9	6	64.4	8.9	x	x	3.0 T S	ICA	DMN, thalamic, ECN
Pasquini et al. [80]*	43	21	8	13	72.3	8.6	22	6	16	66.3	9.0	x	x	3.0 T P	ICA	DMN (a, p)
Adriaanse et al. [28]	59	28	17	11	72.0	4.9	31	17	14	72.0	4.3	x		1.5 T S	ICA	DMN, VIS (med, lat), AN, SMN, ECN, dorsovisual (l, r)
Yi et al. [81]*	23	11	1	10	64.2	2.4	12	3	9	71.8	1.2	x		3.0 T GE	ICA	DMN, SN

(*Continued*)

Table 1
Characteristics of rsfMRI studies included in the meta-analysis (*Continued*)

		MCI					HC					MCI				
Study	*N*	*n*	M	F	Age	SD	*n*	M	F	Age	SD	<HC	>HC	Scanner	Method	Seed region/ICN investigated
Sorg et al. [82]	40	24	13	11	69.3	8.1	16	10	6	68.1	3.8	x		1.5 T S	ICA	VIS, AN, ATN (v), spatial attention, DMN
Bai et al. [83]	56	30	15	15	72.5	4.4	26	12	14	71.6	5.3	x	x	1.5 T GE	SB	PCC
Gili et al. [67]*	20	10	6	4	71.2	4.1	10	7	3	64.1	10.5	x		3.0 T S	*SB* and ICA	PCC, mPFC
Bai et al. [84]	44	26	19	7	71.4	4.3	18	10	8	70.3	4.7		x	1.5 T GE	ICA	DMN
Xie et al. [85]	56	30	19	11	72.6	4.8	26	14	12	70.3	4.8	x		1.5 T GE	SB	Postcentral gyrus (l), hippocampus (l), medialFC (l), middleFC (l), precuneus (l, r), insula (l, r)
Jin et al. [86]	16	8	5	3	60.6	3.2	8	4	4	60.9	8.3	x	x	3.0 T GE	ICA	DMN
Han et al. [87]	80	40	7	33	86.3	4.5	40	15	25	86.3	4.5	x	x	1.5 T GE	SB	PCC
Liang et al. [88]	32	16	6	10	68.5	7.8	16	6	10	67.2	8.4	x	x	3.0 T S	SB	AG (l, r), supramarginal gyrus (l, r), intraparietal sulcus (r)
Hahn et al. [89]†	54	28	14	14	69.5	7.1	26	10	16	65.5	7.8	x		3.0 T P	ICA	DMN (a, p), ATN (d, v), ECN (l, r), SMN, VIS
Myers et al. [90]	35	23	14	9	69.3	7.4	12	5	9	63.8	5.2	x		3.0 T P	ICA	DMN (a, p), ATN (l, r, d), SN, prAN
Koch et al. [91]	40	24	14	10	68.2	8.4	16	7	9	64.8	5.4	x		3.0 T P	ICA	DMN (a, p), ATN (l, r, d), SN, prAN
Pasquini et al. [80]*	44	22	11	11	65.3	8.7	22	6	16	66.3	9.0		x	3.0 T P	ICA	DMN (a, p)
Das et al. [92]	69	30	14	16	71.6	6.8	39	18	21	70.6	9.0	x		3.0 T S	SB	Hippocampal subregions
Gardini et al. [93]	42	21	13	8	70.6	4.7	21	7	14	69.8	6.5		x	3.0 T GE	SB	PCC, mPFC
Yi et al. [81]*	32	20	4	16	71.0		12	3	9	71.8	1.2	x	x	3.0 T GE	ICA	DMN, SN

Abbreviations: a, anterior; ACC, anterior cingulate cortex; AD, Alzheimer's disease; AG, angular gyrus; AN, auditory network; ATN, attentional network; B, Brucker; d, dorsal; dlPFC, dorsolateral prefrontal cortex; DMN, default mode network; ECN, executive control network; F, female; GE, General Electrics; HC, healthy control; ICA, independent component analysis; ICC, isthmus of cingulate cortex; ICN, intrinsic connectivity network; l, left; lat, lateral; M, male; MCI, mild cognitive impairment; med, medial; medialFC, medial frontal cortex; middleFC, middle frontal cortex; mPFC, medial prefrontal cortex; MrD, marginal division; n, number of subjects; p, posterior; P, Philips; PCC, posterior cingulate cortex; prAN, primary auditory network; prVIS, primary visual network; r, right; rsfMRI, resting-state functional magnetic resonance imaging; S, Siemens; saVC, secondary associative visual cortex; SB, seed based; SMN, sensorimotor network; SD, standard deviation; sVIS, secondary visual network; T, Tesla; v, ventral.

NOTE. Data provided in "bold" indicate seven studies using shared cohorts. Coordinates from these seven studies were subsequently pooled under four studies (indicated by superscript letters a, b, and c), under the corresponding earliest publication using the cohort. In column "Method", when both seed-based and ICA rsfMRI methods were used by a study, the method given in "italics" indicates the method associated with reported coordinates. For column "Seed region/ICN investigated", all seed regions and ICNs investigated are listed, irrespective of significant findings.

*Studies reporting significant coordinates for both AD and MCI patients, relative to matched HC.

†Studies investigating both AD dementia and MCI cohorts.

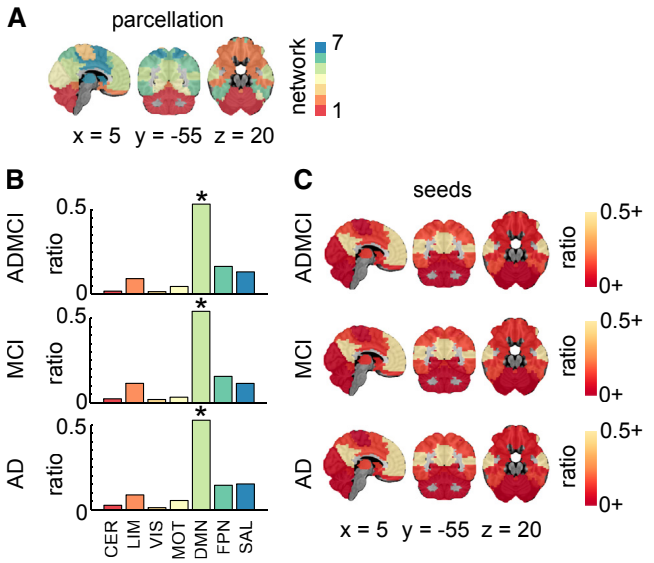

Fig. 2. Seed region network-level findings. (A) R7 atlas; (B) histograms showing the ratio of counts (or hits) across the seven networks for all seeds in ADMCI, MCI, and AD. Significant (* denoting qFDR < 0.05) prevalence of seeds in the DMN was demonstrated across all three cohorts; (C) seed region hit maps (ratio of dysconnectivity coordinates in each network) at R7. Maps are superimposed onto the anatomic International Consortium for Brain Mapping (ICBM) 152 template. x, y, and z Montreal Neurological Institute (MNI) coordinates are given for sagittal, coronal, and axial slices. Abbreviations: AD, Alzheimer's disease; ADMCI, AD dementia and MCI; CER, cerebellar network; DMN, default mode network; FDR, false discovery rate; FPN, frontoparietal network; HC, healthy control; LIM, limbic network; MCI, mild cognitive impairment; MOT, motor network; SAL, salience network; VIS, visual network.

hypoconnectivity and hyperconnectivity in the DMN. Significant hyperconnectivity in the DMN and limbic network (LIM) was observed in the MCI cohort. There was also significant hypoconnectivity in the DMN for the AD group, which appeared as a trend for the MCI group.

We then refined the spatial localization of effects found in R7 using the R36 atlas. Significant DMN hypoconnectivity in AD and ADMCI cohorts was detected in the precuneus (PCu) and posterior cingulate cortex (PCC) (Fig. 4). A trend for DMN hyperconnectivity was observed in the PCu for ADMCI and in both the PCu and PCC in MCI (Fig. 4). The LIM hyperconnectivity was observed as a trend in the hippocampus and entorhinal cortex in MCI patients (Fig. 4).

Finally, we investigated the robustness of findings with respect to the selection of seeds (DMN, non-DMN, or all combined), using the R7 atlas. Significant network-level findings derived from all seeds combined, as reported previously, replicated when using DMN seeds alone (Fig. 3A). In addition, a trend toward hypoconnectivity in MCI became significant using DMN seeds only. When focusing on non-DMN seed studies, no significant effects were observed in the DMN, as expected. The only significant result was hyperconnectivity of the salience network (SAL) in ADMCI, also present as a trend in AD subjects.

3.3. Voxel-based meta-analysis

ALE results demonstrated significant hypoconnectivity in the PCC and PCu in the ADMCI and AD studies (Fig. 5, Supplementary Table 4), consistent with our network-level findings using R7 and R36 atlases. This observation was made both for all seeds combined and DMN-only seeds (Fig. 5, Supplementary Table 4).

Unlike the network-level analysis, using ALE we found diminished connectivity in the primary visual cortex, both in ADMCI and AD. This was observed for all seeds combined as well as for DMN-only seeds in ADMCI and DMN-only seeds in AD. Finally, significant hyperconnectivity was observed in AD in the anterior insula (Fig. 5, Supplementary Table 4), consistent with the trend in the LIM observed using the R36 atlas.

4. Discussion

We report on a systematic meta-analysis of rsfMRI brain connectivity dysfunction in LOAD, using voxel-, region-, and network-level statistics. Our results demonstrated consistent connectivity alterations both within and outside of the DMN.

4.1. Connectivity changes in the DMN

4.1.1. Late-onset AD

Our results revealed a consistent decrease in DMN connectivity in the ADMCI and AD cohorts, particularly in the PCu and PCC, for all resolutions of meta-analysis. This finding is in line with previous meta-analyses centered on the DMN [21,22], and a recent study published after we completed our analysis [23]. DMN deterioration appears robust to the choice of analytical approaches, as previous meta-analyses largely included studies measuring regional homogeneity and amplitude of low-frequency fluctuation. Moreover, our results support previous literature reporting on the vulnerability of the DMN to multiple AD pathophysiology [24].

Unlike our robust findings in AD subjects, DMN hypoconnectivity in MCI could only be demonstrated using network-level statistics, suggesting a weaker, more distributed effect in MCI. However, we recently reported decreased DMN connectivity in a large multisite MCI cohort with a connectome-wide approach [13]. The modest findings of our present meta-analysis may be because of a lack of statistical power from having multiple, small, single-site samples. Clinical heterogeneity might also have played a role, that is, only a subset of MCI patients develop AD dementia [6,25], and there may be pathologic subtypes [26]. We also demonstrated DMN hyperconnectivity in MCI and ADMCI using network-level statistics. These changes may reflect both functional disconnection and compensation in response to damage at earlier stages of neurodegeneration, as well as direct or indirect pathologic mechanisms [27]. Moreover,

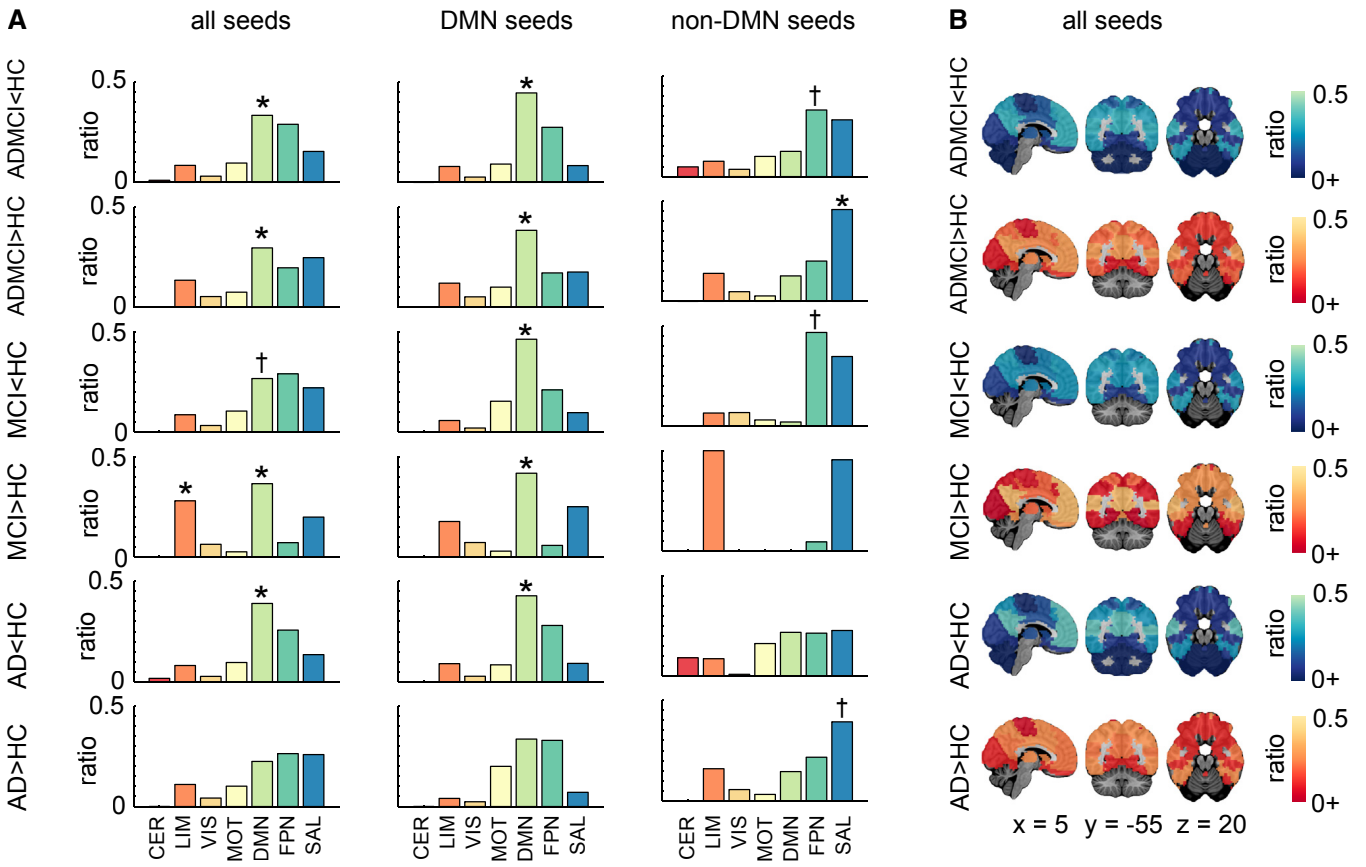

Fig. 3. Network-level findings using the R7 atlas. (A) Histograms showing per contrast the ratio of hits across the seven networks for all seeds, DMN seeds only, and non-DMN seeds. Networks with significant count (or hit) ratios are indicated by * denoting qFDR < 0.05, whereas † denotes $P < .05$ uncorrected. (B) Hit maps at R7 are shown for contrasts ADMCI < HC, ADMCI > HC, MCI < HC, MCI > HC, AD < HC, and AD > HC. Maps are superimposed onto the anatomic ICBM 152 template. x, y, and z MNI coordinates are given for sagittal, coronal, and axial slices. Abbreviations: AD, Alzheimer's disease dementia; ADMCI, AD dementia and MCI; CER, cerebellar network; DMN, default mode network; FDR, false discovery rate; FPN, frontoparietal network; HC, healthy control; LIM, limbic network; MCI, mild cognitive impairment; MOT, motor network; SAL, salience network; VIS, visual network.

there is some uncertainty of the specific nodes that actually show aberrant connectivity in our network-level analysis. This may give rise to apparent contradictory results.

4.1.2. Early onset AD

DMN hypoconnectivity of similar magnitude to LOAD was demonstrated in early onset non–ADAD [28,29], whereas in ADAD, DMN hypoconnectivity was slightly more pronounced than that in LOAD [30]. Altered DMN connectivity was observed in asymptomatic mutation carriers (*PSEN1*, *PSEN2*, or *APP*) many years before the age at which they were expected to develop symptoms [31–33], suggesting that aberrant connectivity may be a very early biomarker for AD.

4.1.3. Cognitively normal individuals at genetic risk for LOAD

Altered DMN connectivity has been reported in cognitively normal APOEε4 carriers compared with non-APOEε4 carriers. These alterations were found across all age groups, that is, elderly [12,34–36], middle-aged [37–39], and young adults [40,41], and were associated with worse cognition in middle-aged and elderly carriers [35,37,39]. Studies have also reported connectivity changes in the DMN in the absence of Pittsburgh compound B–detectable brain amyloidosis [12,40,41], further validating the potential of rsfMRI connectivity as an early marker of synaptic and neuronal dysfunction in AD.

4.1.4. Cognitively normal elderly at risk for LOAD

Aberrant DMN dysconnectivity, particularly reduced connectivity between the anterior and posterior DMN, has been associated with aging and age-related cognitive decline [33,42]. DMN hypoconnectivity may arise as early as middle age [43,44], with decreases occurring at differing rates between sexes [45] most likely due to the differential effect of sex on AD risk [46]. Reduced DMN integrity has also been reported in cognitively normal elderly with abnormal levels of CSF amyloid or tau proteins [47], as well as PET-detectable cerebral amyloidosis [48]. These results suggest that some of the effects related to normal aging in the literature may be driven by preclinical AD. Very few

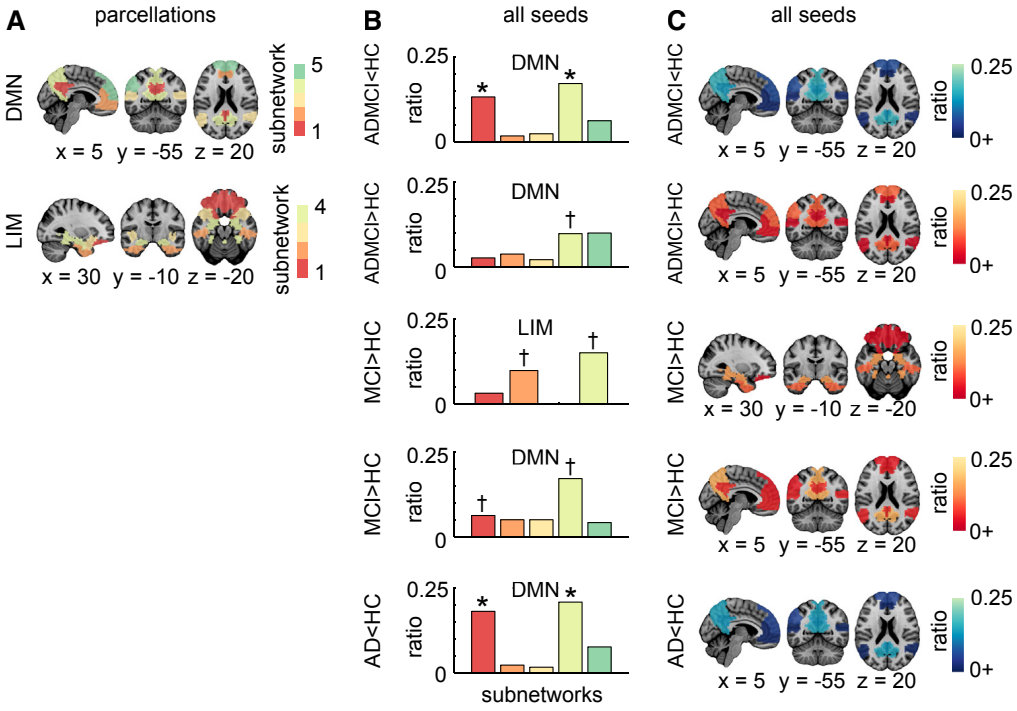

Fig. 4. Network-level findings using the R36 atlas. (A) Functional template at R36 showing the breakdown of the DMN and LIM into subnetworks. These two networks were significant (qFDR < 0.05 for contrasts ADMCI < HC for DMN, ADMCI > HC for DMN, MCI > HC for DMN and LIM, AD < HC for DMN) or trended toward significance ($P < .05$ uncorrected for contrasts MCI < HC for DMN) for the "all-seeds" condition at R7. (B) Histograms showing per selected contrast (as described in A), the ratio of counts (or hits) across the subnetworks. Subnetworks with significant hit ratios are indicated by * denoting qFDR < 0.05, whereas † denotes a trend with $P < .05$ uncorrected. (C) Hit maps at R36 for brain regions that overlap with significant or trending toward significance networks (as described in A). Maps are superimposed onto the anatomic ICBM 152 template. x, y, and z MNI coordinates are given for sagittal, coronal, and axial slices. Abbreviations: AD, Alzheimer's disease dementia; ADMCI, AD dementia and MCI; CER, cerebellar network; DMN, default mode network; FDR, false discovery rate; FPN, frontoparietal network; HC, healthy control; LIM, limbic network; MCI, mild cognitive impairment; MOT, motor network; SAL, salience network; VIS, visual network.

studies examined the interactions between age, sex, LOAD, and rsfMRI connectivity, which is clearly an important avenue for future work.

4.2. Connectivity changes outside the DMN

Our meta-analysis confirmed that intrinsic connectivity disruptions in LOAD are not confined to the DMN. We found increased connectivity in the SAL in ADMCI and AD. Abnormal SAL connectivity has now been reported in another LOAD study [49] published after we completed our meta-analysis and has also been demonstrated in ADAD [30], APOEε4 carriers [36,37], and the elderly [50], with connectivity increases highlighted in APOEε4 carriers. With the anterior insula as a key hub, the SAL plays a pivotal role in network switching between the DMN and frontoparietal network (FPN), two networks exhibiting competitive interactions during cognitive information processing [51]. Association of heightened SAL connectivity with reduced DMN connectivity in AD suggests that progressive DMN impairment may be deleterious to SAL function [52].

We also found increased connectivity in the LIM in MCI. Heightened LIM connectivity has been reported in early onset, non-ADAD patients [29], and in individuals with subjective memory impairment [53]. The effect of APOEε4 carriage on LIM connectivity, however, lacks consensus [54–56]. Since LIM hyperconnectivity in early onset AD patients was shown to correlate positively with memory performance, it is likely that increased connectivity in this network contributes to preserving function in the face of medial temporal lobe pathology [29].

4.3. Selective vulnerability of multimodal networks in AD

The DMN, SAL, and FPN are multimodal networks that interconnect cortical regions associated with various cognitive functions, and they have been demonstrated computationally to support integrative information processing at the cost of being vulnerable to early and fast spreading of insults [57]. Supporting this theoretical finding is the recent observation that tau and amyloid-β, despite their independent patterns of spatial deposition, overlap with brain tissue loss in hub regions of multimodal networks [58]. These multimodal networks are also metabolically expensive and display higher rates of cerebral blood flow, aerobic glycolysis, and oxidative

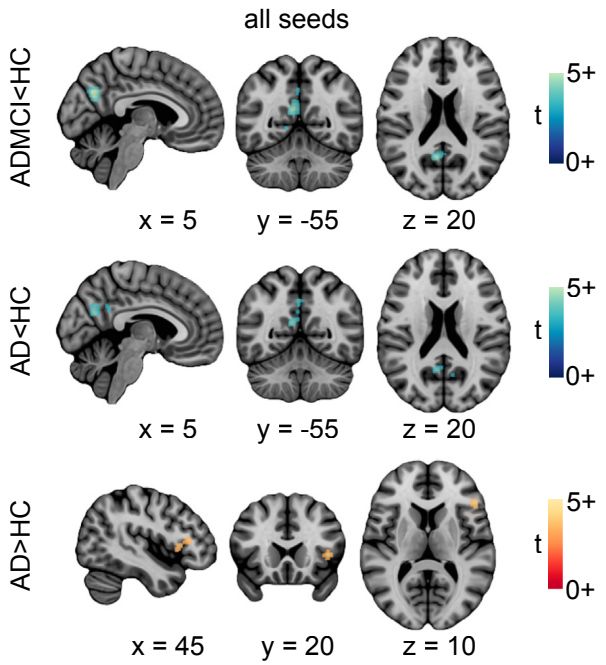

Fig. 5. Location of significant convergence of the voxel-level findings. Regions exhibiting significant rsfMRI abnormalities for contrasts ADMCI < HC, AD < HC, and AD > HC. Activation likelihood estimation images were thresholded at $P < .05$ (cluster-level family wise error or FWE corrected for multiple comparisons; cluster-forming threshold $P < .001$ at the voxel level) and displayed as t scores, with hyperconnectivity in red-orange and hypoconnectivity in blue-green. Maps are superimposed onto the anatomic ICBM 152 template. x, y, and z MNI coordinates are given for sagittal, coronal, and axial slices. Abbreviations: AD, Alzheimer's disease dementia; ADMCI, AD dementia and mild cognitive impairment; HC, healthy control; rsfMRI, resting-state functional magnetic resonance imaging.

glucose metabolism [59]. The high-value/high-cost characteristics of the DMN, SAL, and FPN may make them vulnerable to AD-associated pathogenic processes, such as metabolic dysfunction/oxidative stress, and accumulation of toxic proteins, such as amyloid-β [59]. The hypothesis that multimodal networks/regions are particularly susceptible to AD-associated pathophysiological processes may explain our finding of consistent alterations of these networks.

4.4. Limitations

Our literature search did not identify an abundance of rsfMRI literature in AD and MCI cohorts, which clearly expresses the need for additional research. The relatively low number of experiments that met our inclusion criteria might have underpowered our voxel-level findings, especially for the MCI contrasts. In addition, our search demonstrated that typical studies featured small samples, and also that analytical methods were quite variable in the field (a main reason for excluding an article was due to methodology used). This setting is particularly amenable to questionable research practices, including "p-hacking" (testing several methods, reporting only one). Given the near absence of negative results reporting in the field, on one hand, and the large size of the rsfMRI field, on the other hand, there is no question that some amount of publication bias is also present. Meta-analytical tools, such as funnel plots, are available to detect both selective reporting and p-hacking but are not feasible given current reporting practices in the rsfMRI community [60].

Another limitation of our study is experimental heterogeneity, in terms of population recruitment, scan acquisition (e.g., scanner make and model, scanning parameters), and processing choices [13,20,61]. The prominence of the DMN in our results partly reflects the focus on this network in the literature, which we quantified using seed statistics. Hypothesis-driven analyses on the DMN are attractive for assessing connectivity changes in small samples; as such analyses will have good statistical power if the DMN truly carries the larger effects in the brain. However, full-brain studies will be required to get a more comprehensive view on AD-related changes in rsfMRI network connectivity using meta-analyses. The current trend toward large public samples [13,62] is enabling unbiased meta-analyses, pooling neuroimaging data across many studies instead of relying on published coordinates. This will hopefully resolve most of the aforementioned limitations in the future.

5. Conclusions

Our meta-analysis demonstrated consistent connectivity alterations in the DMN, SAL, and LIM in the spectrum of LOAD, supporting the use of resting-state connectivity as a biomarker of AD.

Acknowledgments

The computational resources used to perform the data analysis were provided by Compute Canada (www.computecanada.org) and CLUMEQ (www.clumeq.mcgill.ca), which is funded in part by NSERC (MRS), FQRNT, and McGill University.

This research was supported by the Canadian Consortium on Neurodegeneration in Aging. The Canadian Consortium on Neurodegeneration in Aging is supported by a grant from the Canadian Institutes of Health Research with funding from several partners including the Alzheimer Society of Canada, Sanofi, and Women's Brain Health Initiative. This research was also supported by the Courtois Foundation (P.B.) and an Alzheimer Society Postdoctoral Fellowship (A.B.).

Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.dadm.2017.03.007.

RESEARCH IN CONTEXT

1. Systematic review: We conducted a systematic review of PubMed-indexed resting-state functional magnetic resonance imaging (rsfMRI) studies in accordance with the "Preferred Reporting Items for Systematic Reviews and Meta-Analyses" guidelines. We included studies that investigated differences in functional connectivity, relative to controls, between patients with Alzheimer's disease (AD) and/or mild cognitive impairment, and reported coordinates of findings.
2. Interpretation: Typical rsfMRI functional connectivity studies in AD suffer from low statistical power. Our meta-analysis quantifies if and where convergent findings have been reported in the literature and strengthens the evidence for the use of rsfMRI as an AD biomarker.
3. Future directions: A disproportionately large portion of studies specifically investigated the default mode network, based on well-grounded hypotheses on AD pathophysiology. It is unclear if AD truly has larger effects on default-mode connectivity because of limited power to examine other networks. Future research should aim for full-brain investigations using larger study populations.

References

[1] Sperling RA, Aisen PS, Beckett LA, Bennett DA, Craft S, Fagan AM, et al. Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011; 7:280–92.

[2] Wu L, Rosa-Neto P, Hsiung GY, Sadovnick AD, Masellis M, Black SE, et al. Early-onset familial Alzheimer's disease (EOFAD). Can J Neurol Sci 2012;39:436–45.

[3] Campion D, Dumanchin C, Hannequin D, Dubois B, Belliard S, Puel M, et al. Early-onset autosomal dominant Alzheimer disease: prevalence, genetic heterogeneity, and mutation spectrum. Am J Hum Genet 1999;65:664–70.

[4] Hughes RE, Nikolic K, Ramsay RR. One for all? Hitting multiple Alzheimer's disease targets with one drug. Front Neurosci 2016;10:177.

[5] Matthews PM, Hampshire A. Clinical concepts emerging from fMRI functional connectomics. Neuron 2016;91:511–28.

[6] Albert MS, DeKosky ST, Dickson D, Dubois B, Feldman HH, Fox NC, et al. The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011; 7:270–9.

[7] Vemuri P, Jones DT, Jack CR Jr. Resting state functional MRI in Alzheimer's disease. Alzheimers Res Ther 2012;4:2.

[8] Logothetis NK. The neural basis of the blood-oxygen-level-dependent functional magnetic resonance imaging signal. Philos Trans R Soc Lond B Biol Sci 2002;357:1003–37.

[9] Krajcovicova L, Marecek R, Mikl M, Rektorova I. Disruption of resting functional connectivity in Alzheimer's patients and at-risk subjects. Curr Neurol Neurosci Rep 2014;14:491.

[10] Brier MR, Thomas JB, Ances BM. Network dysfunction in Alzheimer's disease: refining the disconnection hypothesis. Brain Connect 2014;4:299–311.

[11] Jack CR Jr, Knopman DS, Jagust WJ, Petersen RC, Weiner MW, Aisen PS, et al. Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers. Lancet Neurol 2013;12:207–16.

[12] Sheline YI, Morris JC, Snyder AZ, Price JL, Yan Z, D'Angelo G, et al. APOE4 allele disrupts resting state fMRI connectivity in the absence of amyloid plaques or decreased CSF Aβ42. J Neurosci 2010; 30:17035–40.

[13] Tam A, Dansereau C, Badhwar A, Orban P, Belleville S, Chertkow H, et al. Common effects of amnestic mild cognitive impairment on resting-state connectivity across four independent studies. Front Aging Neurosci 2015;7:242.

[14] Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Int J Surg 2010;8:336–41.

[15] Bellec P, Benhajali Y, Carbonell F, Dansereau C, Albouy G, Pelland M, et al. Impact of the resolution of brain parcels on connectome-wide association studies in fMRI. Neuroimage 2015;123:212–28.

[16] Lancaster JL, Tordesillas-Gutiérrez D, Martinez M, Salinas F, Evans A, Zilles K, et al. Bias between MNI and Talairach coordinates analyzed using the ICBM-152 brain template. Hum Brain Mapp 2007; 28:1194–205.

[17] Phipson B, Smyth GK. Permutation P-values should never be zero: calculating exact P-values when permutations are randomly drawn. Stat Appl Genet Mol Biol 2010;9. Article39.

[18] Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B Stat Methodol 1995;57:289–300.

[19] Eickhoff SB, Bzdok D, Laird AR, Kurth F, Fox PT. Activation likelihood estimation meta-analysis revisited. Neuroimage 2012; 59:2349–61.

[20] Eickhoff SB, Nichols TE, Laird AR, Hoffstaedter F, Amunts K, Fox PT, et al. Behavior, sensitivity, and power of activation likelihood estimation characterized by massive empirical simulation. Neuroimage 2016;137:70–85.

[21] Li HJ, Hou XH, Liu HH, Yue CL, He Y, Zuo XN. Toward systems neuroscience in mild cognitive impairment and Alzheimer's disease: a meta-analysis of 75 fMRI studies. Hum Brain Mapp 2015; 36:1217–32.

[22] Jacobs HIL, Radua J, Lückmann HC, Sack AT. Meta-analysis of functional network alterations in Alzheimer's disease: toward a network biomarker. Neurosci Biobehav Rev 2013;37:753–65.

[23] Kim HJ, Cha J, Lee JM, Shin JS, Jung NY, Kim YJ, et al. Distinctive resting state network disruptions among Alzheimer's disease, subcortical vascular dementia, and mixed dementia patients. J Alzheimers Dis 2016;50:709–18.

[24] Buckner RL, Sepulcre J, Talukdar T, Krienen FM, Liu H, Hedden T, et al. Cortical hubs revealed by intrinsic functional connectivity: mapping, assessment of stability, and relation to Alzheimer's disease. J Neurosci 2009;29:1860–73.

[25] Malek-Ahmadi M. Reversion from mild cognitive impairment to normal cognition: a meta-analysis. Alzheimer Dis Assoc Disord 2016;30:324–30.

[26] Köhler S, Hamel R, Sistermans N, Koene T, Pijnenburg YAL, van der Flier WM, et al. Progression to dementia in memory clinic patients without dementia: a latent profile analysis. Neurology 2013;81:1342–9.

[27] Jones DT, Knopman DS, Gunter JL, Graff-Radford J, Vemuri P, Boeve BF, et al. Cascading network failure across the Alzheimer's disease spectrum. Brain 2016;139:547–62.

[28] Adriaanse SM, Binnewijzend MA, Ossenkoppele R, Tijms BM, van der Flier WM, Koene T, et al. Widespread disruption of functional

brain organization in early-onset Alzheimer's disease. PLoS One 2014;9:e102995.
[29] Gour N, Felician O, Didic M, Koric L, Gueriot C, Chanoine V, et al. Functional connectivity changes differ in early and late-onset Alzheimer's disease. Hum Brain Mapp 2014;35:2978–94.
[30] Thomas JB, Brier MR, Bateman RJ, Snyder AZ, Benzinger TL, Xiong C, et al. Functional connectivity in autosomal dominant and late-onset Alzheimer disease. JAMA Neurol 2014;71:1111–22.
[31] Quiroz YT, Schultz AP, Chen K, Protas HD, Brickhouse M, Fleisher AS, et al. Brain imaging and blood biomarker abnormalities in children with autosomal dominant Alzheimer disease: a cross-sectional study. JAMA Neurol 2015;72:912–9.
[32] Chhatwal JP, Schultz AP, Johnson K, Benzinger TL, Jack C Jr, Ances BM, et al. Impaired default network functional connectivity in autosomal dominant Alzheimer disease. Neurology 2013;81:736–44.
[33] Sala-Llonch R, Fortea J, Bartrés-Faz D, Bosch B, Lladó A, Peña-Gómez C, et al. Evolving brain functional abnormalities in PSEN1 mutation carriers: a resting and visual encoding fMRI study. J Alzheimers Dis 2013;36:165–75.
[34] Shu H, Shi Y, Chen G, Wang Z, Liu D, Yue C, et al. Opposite neural trajectories of apolipoprotein E ε4 and ε2 alleles with aging associated with different risks of Alzheimer's disease. Cereb Cortex 2016;26:1421–9.
[35] Song H, Long H, Zuo X, Yu C, Liu B, Wang Z, et al. APOE effects on default mode network in Chinese cognitive normal elderly: relationship with clinical cognitive performance. PLoS One 2015;10:e0133179.
[36] Machulda MM, Jones DT, Vemuri P, McDade E, Avula R, Przybelski S, et al. Effect of APOE ε4 status on intrinsic network connectivity in cognitively normal elderly subjects. Arch Neurol 2011;68:1131–6.
[37] Goveas JS, Xie C, Chen G, Li W, Ward BD, Franczak MB, et al. Functional network endophenotypes unravel the effects of apolipoprotein E epsilon 4 in middle-aged adults. PLoS One 2013;8:e55902.
[38] Patel KT, Stevens MC, Pearlson GD, Winkler AM, Hawkins KA, Skudlarski P, et al. Default mode network activity and white matter integrity in healthy middle-aged ApoE4 carriers. Brain Imaging Behav 2013;7:60–7.
[39] Westlye ET, Lundervold A, Rootwelt H, Lundervold AJ, Westlye LT. Increased hippocampal default mode synchronization during rest in middle-aged and elderly APOE ε4 carriers: relationships with memory performance. J Neurosci 2011;31:7775–83.
[40] Su YY, Liang X, Schoepf UJ, Varga-Szemes A, West HC, Qi R, et al. APOE polymorphism affects brain default mode network in healthy young adults: a STROBE article. Medicine 2015;94:e1734.
[41] Filippini N, MacIntosh BJ, Hough MG, Goodwin GM, Frisoni GB, Smith SM, et al. Distinct patterns of brain activity in young carriers of the APOE-epsilon4 allele. Proc Natl Acad Sci U S A 2009;106:7209–14.
[42] Ferreira LK, Busatto GF. Resting-state functional connectivity in normal brain aging. Neurosci Biobehav Rev 2013;37:384–400.
[43] Evers EA, Klaassen EB, Rombouts SA, Backes WH, Jolles J. The effects of sustained cognitive task performance on subsequent resting state functional connectivity in healthy young and middle-aged male school teachers. Brain Connect 2012;2:102–12.
[44] Zuo XN, Kelly C, Di Martino A, Mennes M, Margulies DS, Bangaru S, et al. Growing together and growing apart: regional and sex differences in the lifespan developmental trajectories of functional homotopy. J Neurosci 2010;30:15034–43.
[45] Scheinost D, Finn ES, Tokoglu F, Shen X, Papademetris X, Hampson M, et al. Sex differences in normal age trajectories of functional brain networks. Hum Brain Mapp 2015;36:1524–35.
[46] Podcasy JL, Epperson CN. Considering sex and gender in Alzheimer disease and other dementias. Dialogues Clin Neurosci 2016;18:437–46.
[47] Wang L, Brier MR, Snyder AZ, Thomas JB, Fagan AM, Xiong C, et al. Cerebrospinal fluid Aβ42, phosphorylated Tau181, and resting-state functional connectivity. JAMA Neurol 2013;70:1242–8.
[48] Elman JA, Madison CM, Baker SL, Vogel JW, Marks SM, Crowley S, et al. Effects of beta-amyloid on resting state functional connectivity within and between networks reflect known patterns of regional vulnerability. Cereb Cortex 2016;26:695–707.
[49] Wang Z, Zhang M, Han Y, Song H, Guo R, Li K. Differentially disrupted functional connectivity of the subregions of the amygdala in Alzheimer's disease. J Xray Sci Technol 2016;24:329–42.
[50] Sala-Llonch R, Bartrés-Faz D, Junqué C. Reorganization of brain networks in aging: a review of functional connectivity studies. Front Psychol 2015;6:663.
[51] He X, Qin W, Liu Y, Zhang X, Duan Y, Song J, et al. Abnormal salience network in normal aging and in amnestic mild cognitive impairment and Alzheimer's disease. Hum Brain Mapp 2014;35:3446–64.
[52] Zhou J, Greicius MD, Gennatas ED, Growdon ME, Jang JY, Rabinovici GD, et al. Divergent network connectivity changes in behavioural variant frontotemporal dementia and Alzheimer's disease. Brain 2010;133:1352–67.
[53] Gour N, Ranjeva JP, Ceccaldi M, Confort-Gouny S, Barbeau E, Soulier E, et al. Basal functional connectivity within the anterior temporal network is associated with performance on declarative memory tasks. Neuroimage 2011;58:687–97.
[54] Heise V, Filippini N, Trachtenberg AJ, Suri S, Ebmeier KP, Mackay CE. Apolipoprotein E genotype, gender and age modulate connectivity of the hippocampus in healthy adults. Neuroimage 2014;98:23–30.
[55] Matura S, Prvulovic D, Butz M, Hartmann D, Sepanski B, Linnemann K, et al. Recognition memory is associated with altered resting-state functional connectivity in people at genetic risk for Alzheimer's disease. Eur J Neurosci 2014;40:3128–35.
[56] Trachtenberg AJ, Filippini N, Ebmeier KP, Smith SM, Karpe F, Mackay CE. The effects of APOE on the functional architecture of the resting brain. Neuroimage 2012;59:565–72.
[57] Mišić B, Betzel RF, Nematzadeh A, Goñi J, Griffa A, Hagmann P, et al. Cooperative and competitive spreading dynamics on the human connectome. Neuron 2015;86:1518–29.
[58] Sepulcre J, Schultz AP, Sabuncu M, Gomez-Isla T, Chhatwal J, Becker A, et al. In vivo tau, amyloid, and gray matter profiles in the aging brain. J Neurosci 2016;36:7364–74.
[59] Crossley NA, Mechelli A, Scott J, Carletti F, Fox PT, McGuire P, et al. The hubs of the human connectome are generally implicated in the anatomy of brain disorders. Brain 2014;137:2382–95.
[60] Sterne JA, Egger M. Funnel plots for detecting bias in meta-analysis: guidelines on choice of axis. J Clin Epidemiol 2001;54:1046–55.
[61] Jones DT, Vemuri P, Murphy MC, Gunter JL, Senjem ML, Machulda MM, et al. Non-stationarity in the "resting brain's" modular architecture. PLoS One 2012;7:e39731.
[62] Cheng W, Palaniyappan L, Li M, Kendrick KM, Zhang J, Luo Q, et al. Voxel-based, brain-wide association study of aberrant functional connectivity in schizophrenia implicates thalamocortical circuitry. NPJ Schizophr 2015;1:15016.
[63] Wang K, Liang M, Wang L, Tian L, Zhang X, Li K, et al. Altered functional connectivity in early Alzheimer's disease: a resting-state fMRI study. Hum Brain Mapp 2007;28:967–78.
[64] Zhang H-Y, Wang S-J, Xing J, Liu B, Ma Z-L, Yang M, et al. Detection of PCC functional connectivity characteristics in resting-state fMRI in mild Alzheimer's disease. Behav Brain Res 2009;197:103–8.
[65] Zhang H-Y, Wang S-J, Liu B, Ma Z-L, Yang M, Zhang Z-J, et al. Resting brain connectivity: changes during the progress of Alzheimer disease. Radiology 2010;256:598–606.
[66] Sheline YI, Raichle ME, Snyder AZ, Morris JC, Head D, Wang S, et al. Amyloid plaques disrupt resting state default mode network connectivity in cognitively normal elderly. Biol Psychiatry 2010;67:584–7.

[67] Gili T, Cercignani M, Serra L, Perri R, Giove F, Maraviglia B, et al. Regional brain atrophy and functional disconnection across Alzheimer's disease evolution. J Neurol Neurosurg Psychiatry 2011;82:58–66.
[68] Wu X, Li R, Fleisher AS, Reiman EM, Guan X, Zhang Y, et al. Altered default mode network connectivity in Alzheimer's disease–a resting functional MRI and Bayesian network study. Hum Brain Mapp 2011;32:1868–81.
[69] Li R, Wu X, Fleisher AS, Reiman EM, Chen K, Yao L. Attention-related networks in Alzheimer's disease: a resting functional MRI study. Hum Brain Mapp 2012;33:1076–88.
[70] Damoiseaux JS, Prater KE, Miller BL, Greicius MD. Functional connectivity tracks clinical deterioration in Alzheimer's disease. Neurobiol Aging 2012;33:828.e19–30.
[71] Binnewijzend MAA, Schoonheim MM, Sanz-Arigita E, Wink AM, van der Flier WM, Tolboom N, et al. Resting-state fMRI changes in Alzheimer's disease and mild cognitive impairment. Neurobiol Aging 2012;33:2018–28.
[72] Kenny ER, Blamire AM, Firbank MJ, O'Brien JT. Functional connectivity in cortical regions in dementia with Lewy bodies and Alzheimer's disease. Brain 2012;135:569–81.
[73] Zhu DC, Majumdar S, Korolev IO, Berger KL, Bozoki AC. Alzheimer's disease and amnestic mild cognitive impairment weaken connections within the default-mode network: a multi-modal imaging study. J Alzheimers Dis 2013;34:969–84.
[74] Balthazar MLF, Pereira FRS, Lopes TM, da Silva EL, Coan AC, Campos BM, et al. Neuropsychiatric symptoms in Alzheimer's disease are related to functional connectivity alterations in the salience network. Hum Brain Mapp 2014;35:1237–46.
[75] Yao H, Liu Y, Zhou B, Zhang Z, An N, Wang P, et al. Decreased functional connectivity of the amygdala in Alzheimer's disease revealed by resting-state fMRI. Eur J Radiol 2013;82:1531–8.
[76] Zhou B, Liu Y, Zhang Z, An N, Yao H, Wang P, et al. Impaired functional connectivity of the thalamus in Alzheimer's disease and mild cognitive impairment: a resting-state fMRI study. Curr Alzheimer Res 2013;10:754–66.
[77] Zhang Z, Liu Y, Zhou B, Zheng J, Yao H, An N, et al. Altered functional connectivity of the marginal division in Alzheimer's disease. Curr Alzheimer Res 2014;11:145–55.
[78] Weiler M, Fukuda A, Massabki LHP, Lopes TM, Franco AR, Damasceno BP, et al. Default mode, executive function, and language functional connectivity networks are compromised in mild Alzheimer's disease. Curr Alzheimer Res 2014;11:274–82.
[79] Balachandar R, John JP, Saini J, Kumar KJ, Joshi H, Sadanand S, et al. A study of structural and functional connectivity in early Alzheimer's disease using rest fMRI and diffusion tensor imaging. Int J Geriatr Psychiatry 2015;30:497–504.
[80] Pasquini L, Scherr M, Tahmasian M, Meng C, Myers NE, Ortner M, et al. Link between hippocampus' raised local and eased global intrinsic connectivity in AD. Alzheimers Dement 2015;11:475–84.
[81] Yi D, Choe YM, Byun MS, Sohn BK, Seo EH, Han J, et al. Differences in functional brain connectivity alterations associated with cerebral amyloid deposition in amnestic mild cognitive impairment. Front Aging Neurosci 2015;7:15.
[82] Sorg C, Riedl V, Mühlau M, Calhoun VD, Eichele T, Läer L, et al. Selective changes of resting-state networks in individuals at risk for Alzheimer's disease. Proc Natl Acad Sci U S A 2007;104:18760–5.
[83] Bai F, Watson DR, Yu H, Shi Y, Yuan Y, Zhang Z. Abnormal resting-state functional connectivity of posterior cingulate cortex in amnestic type mild cognitive impairment. Brain Res 2009;1302:167–74.
[84] Bai F, Watson DR, Shi Y, Wang Y, Yue C, YuhuanTeng, et al. Specifically Progressive Deficits of Brain Functional Marker in Amnestic Type Mild Cognitive Impairment. PLoS One 2011;6:e24271.
[85] Xie C, Bai F, Yuan B, Yu H, Shi Y, Yuan Y, et al. Joint effects of gray matter atrophy and altered functional connectivity on cognitive deficits in amnestic mild cognitive impairment patients. Psychol Med 2015;45:1799–810.
[86] Jin M, Pelak VS, Cordes D. Aberrant default mode network in subjects with amnestic mild cognitive impairment using resting-state functional MRI. Magn Reson Imaging 2012;30:48–61.
[87] Han SD, Arfanakis K, Fleischman DA, Leurgans SE, Tuminello ER, Edmonds EC, et al. Functional connectivity variations in mild cognitive impairment: associations with cognitive function. J Int Neuropsychol Soc 2012;18:39–48.
[88] Liang P, Wang Z, Yang Y, Li K. Three subsystems of the inferior parietal cortex are differently affected in mild cognitive impairment. J Alzheimers Dis 2012;30:475–87.
[89] Hahn K, Myers N, Prigarin S, Rodenacker K, Kurz A, Förstl H, et al. Selectively and progressively disrupted structural connectivity of functional brain networks in Alzheimer's disease - revealed by a novel framework to analyze edge distributions of networks detecting disruptions with strong statistical evidence. Neuroimage 2013;81:96–109.
[90] Myers N, Pasquini L, Göttler J, Grimmer T, Koch K, Ortner M, et al. Within-patient correspondence of amyloid-β and intrinsic network connectivity in Alzheimer's disease. Brain 2014;137:2052–64.
[91] Koch K, Myers NE, Göttler J, Pasquini L, Grimmer T, Förster S, et al. Disrupted Intrinsic Networks Link Amyloid-β Pathology and Impaired Cognition in Prodromal Alzheimer's Disease. Cereb Cortex 2015;25:4678–88.
[92] Das SR, Pluta J, Mancuso L, Kliot D, Yushkevich PA, Wolk DA. Anterior and posterior MTL networks in aging and MCI. Neurobiol Aging 2015;36 Suppl 1:S141–50.e1.
[93] Gardini S, Venneri A, Sambataro F, Cuetos F, Fasano F, Marchi M, et al. Increased functional connectivity in the default mode network in mild cognitive impairment: a maladaptive compensatory mechanism associated with poor semantic memory performance. J Alzheimers Dis 2015;45:457–70.

Appendix III

Third Appendix

Subtypes of functional brain connectivity as early markers of neurodegeneration in Alzheimer's disease

Pierre Orban, Angela Tam, Melissa Savard, Cécile Madjar, AmanPreet Badhwar, Christian Dansereau, Jacob Vogel, Sylvia Villeneuve, Judes Poirier, Pedro Rosa-Neto, John Breitner, Pierre Bellec, for the Alzheimer's Disease Neuroimaging Initiative* and the Pre-symptomatic Evaluation of Novel or Experimental Treatments for Alzheimer's Disease Program**

* Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

**

Highlights (4 max, 80 characters)

- Functional brain network subtypes associated with cognitive impairment in AD
- Symptom-related subtypes found in the default-mode, limbic and salience networks
- A limbic subtype was associated with familial risk of AD in healthy older adults
- Limbic subtypes associated with beta amyloid deposition and ApoE4

In Brief (40 words)

Orban et al. characterized the heterogeneity of functional connectivity networks in older adults and found network subtypes associated with AD-related clinical symptoms in patients, as well as associations with several AD biomarkers and risk factors in asymptomatic individuals.

Summary (150 words)

Brain degeneration is heterogeneous across individuals, yet this topic has not yet been investigated for functional brain networks, a promising biomarker of Alzheimer's disease. In this study, resting-state functional magnetic resonance imaging was coupled with cluster analysis to capture connectivity subtypes in older adults. Subtypes were first identified in a mixed sample comprised of healthy controls and patients with mild cognitive impairment or Alzheimer's dementia, and associations with symptoms were found in the default-mode, limbic and salience networks. A limbic subtype was then found to be over-represented in an independent, asymptomatic cohort at familial risk of Alzheimer's disease. Other limbic subtypes showed associations with known biomarkers or risk factors for Alzheimer's disease such as cerebrospinal fluid $A\beta_{1-42}$ levels. Our results demonstrate the heterogeneity of functional brain network organization in older adults, and support future investigations in subtypes of functional connectivity in the limbic network as early biomarkers of Alzheimer's degeneration.

Introduction

Alzheimer's disease (AD) is a chronic neurodegenerative disorder that gives rise to the most common form of dementia, manifesting with severe memory and cognitive impairments. However, the clinical expression of AD only becomes apparent decades after the development of neuropathological processes, such as the accumulation of amyloid beta (Aβ) plaques and tau neurofibrillary tangles in the brain. The long preclinical buildup of AD pathology opens the opportunity to prevent, rather than repair, neurodegeneration (Dubois et al., 2016; Sperling et al., 2012). Functional brain connectivity measured with resting-state functional magnetic resonance imaging (rs-fMRI) may be able to capture early synaptic dysfunction in AD (Selkoe, 2002; Tampellini, 2015) and is thus emerging as a very promising biomarker candidate for the diagnosis of AD at the clinical and preclinical stages (Badhwar et al., 2017; Brier et al., 2014; Jones et al., 2016; Vemuri et al., 2012). However, the current literature has largely relied on comparisons between patients and cognitively healthy individuals. Such cross-sectional analyses neglect the considerable phenotypic heterogeneity present both in patient and control populations. The primary objective of this work was to characterize this heterogeneity of functional brain connectivity, and identify network subtypes associated with AD at the clinical and preclinical stages.

A prevalent model of AD postulates that symptoms arise as a consequence of disruptions in distributed networks, rather than local, circumscribed alteration in neural processing (Delbeuck et al., 2003; Seeley et al., 2009). The seminal work of (Greicius et al., 2004) in symptomatic AD demonstrated alterations in functional brain connectivity in the so-called default-mode network (DMN), whose topography overlaps substantially with patterns of end-stage Aβ deposition (Buckner et al., 2005). A recent meta-analysis of over 30 publications looking at functional brain connectivity in AD confirmed the DMN as a key affected brain component (Badhwar et al., 2017). Connectivity disturbances in other large-scale brain networks were also consistently observed in AD, in particular in the limbic and salience networks. At a preclinical stage, current evidence includes a series of associations with well-studied biomarkers or risk factors of AD, especially in the DMN. Rs-fMRI connectivity has been shown to be impacted in cognitively healthy elders at risk of AD due to abnormal levels of cerebrospinal fluid (CSF) $A\beta_{1-42}$ or tau proteins (Jiang et al., 2016; Wang et al., 2013), increased cerebral Aβ deposits (Elman et al., 2016), and presence of apolipoprotein E ε4 allele - ApoE4 (Sheline et al., 2010), the major genetic risk factor in sporadic AD. A familial history of sporadic AD in first-degree relatives is the second most important risk factor of AD (Tanzi, 2012), and was shown to impact on DMN connectivity even in ApoE4 non carriers, thus highlighting additional genetic risk factors (Wang et al., 2012).

Despite mounting evidence of rs-fMRI being a sensitive early marker of AD, identifying the imprint of AD on functional brain connectivity remains challenging due to the heterogeneity present in patients and controls recruited in clinical trials. Post-mortem histological examination of AD pathology in brain tissue samples (Hyman et al., 2012) often does not align with clinical diagnosis (Beach et al., 2012). Over 50% of patients diagnosed with AD dementia in fact do not present Alzheimer’s pathology at a high level

of neuropathological confidence. Conversely, close to 40% of patients diagnosed with non-AD dementia show minimal signs of AD pathology. It is in addition expected that some cognitively healthy persons included in control groups may suffer from preclinical AD, with 10% to 30% of them having Aβ deposition in their brain (Chételat et al., 2013), and some of them exhibiting high loads of neurofibrillary tangles (Mufson et al., 2016). Heterogeneity is also reflected in severity profiles of memory, language, visuospatial and executive impairments in AD, allowing stratification of patients into distinct cognitive subtypes (Scheltens et al., 2016). Data-driven analysis of structural MRI in AD further showed that symptomatic heterogeneity is at least partly related to different modes of atrophy spreading in AD (Dong et al., 2017; Zhang et al., 2016). Separate atrophy subtypes were indeed associated with specific cognitive and clinical profiles, biomarkers, and longitudinal trajectories. Recently, (Doan et al., 2017) also reported various subtypes of dysconnectivity in patients suffering from AD dementia, MCI and subjective cognitive impairment (SCI), using diffusion magnetic resonance imaging, and reported associations between subtypes and the severity of cognitive impairment. The established heterogeneity in structural brain degeneration calls for a data-driven identification of functional connectivity subtypes in older adults.

The overarching goal of the present work was to identify one or multiple subtypes of functional brain connectivity associated with AD, either at a clinical or preclinical stage, using data-driven techniques. We first applied a cluster analysis to identify subgroups of subjects with homogeneous subtypes of brain connectivity within a cohort of 130 subjects, the ADNI2-MTL sample. This mixed sample included patients with dementia of the AD type (hereafter referred to as AD subjects, N = 21), patients with mild cognitive impairment (MCI subjects, N= 44), and elderly healthy controls (HC subjects, N= 65) (Figure 1, Table 1). For each brain network and connectivity subtype, we tested whether

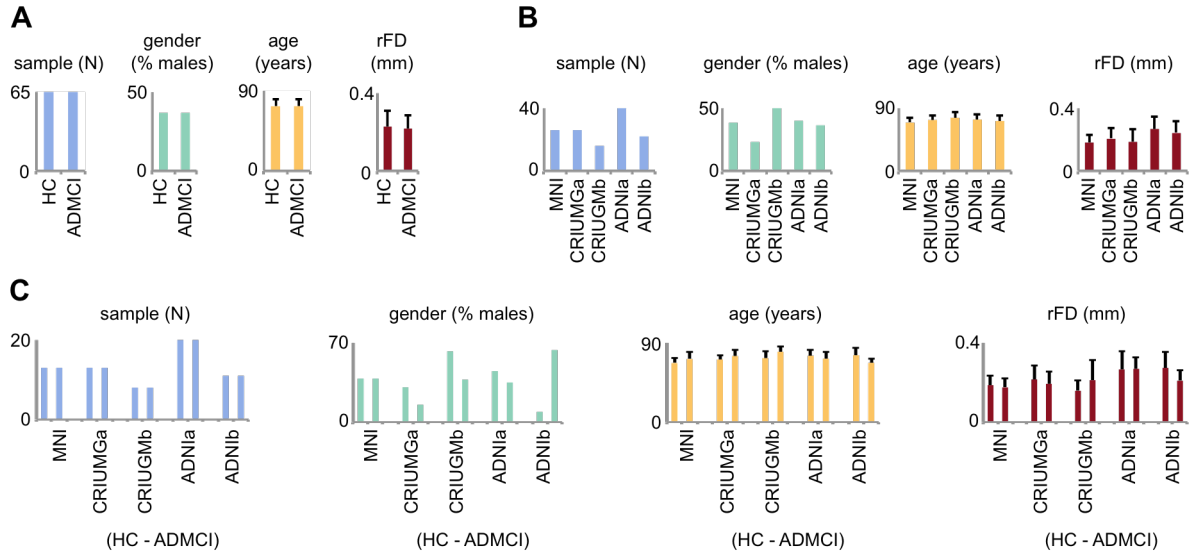

Figure 1. Matching between ADMCI patients and HC
(A) Patients and controls were matched with respect to sample size, gender, age and motion levels after scrubbing (residual frame displacement, rFD). (B) Between-site differences on such variables are shown irrespective of clinical status. (C) The number of patients and controls are perfectly balanced within sites.

a particular subtype was associated with the presence of mild or severe symptoms. AD and MCI subjects were pooled together into a single ADMCI clinical group for these tests as we wished to identify AD-related subtypes largely invariant to disease stage. We then investigated the possibility that some subtypes that were positively associated with symptoms could already be detected at the preclinical stage in a sample of 231 cognitively healthy elders with a familial history of AD (FH subjects). We further investigated the presence of AD-related subtypes in the same cohort of FH subjects by testing their association with known biomarkers or risk factors of AD, namely CSF $A\beta_{1-42}$ and Tau levels as well as ApoE4 genotype.

Results

Subtypes of functional brain networks

To identify subtypes of functional brain networks, we first generated individual functional connectivity maps for seven large-scale networks together covering the entire brain (Figure 2A-B). These reference networks were obtained from an independent dataset and were labeled as cerebellar, limbic, motor, visual, default-mode, fronto-parietal and salience networks. For each network, a hierarchical cluster analysis was applied on 130 individual network maps from the ADNI2-MTL dataset, after regression of phenotypic and site confounds, in order to identify subgroups of subjects with homogeneous brain maps. Visual inspection suggested the presence of at least three voxelwise connectivity subgroups (Figure 2C-D). A brain map averaged across all subjects within a subgroup defined a subtype of network connectivity, highlighting specific brain areas that differed between that subgroup and the overall population average (Figure 2E). Subtypes maps revealed high connectivity with their reference network, yet also exhibited noticeable variations. These differences were not only observed in the associated network (within-network connectivity) but also in other brain areas (between-network connectivity). For instance, subtypes of the DMN could be distinguished from one another not only in terms of connectivity levels within the precuneus or anterior medial prefrontal cortex, two key nodes of the default-mode, but also with regards to connectivity strength in the anterior cingulate, associated with the salience network. For each network, we generated the spatial correlations between individual connectivity maps and each average subtype map, hereafter referred to as weights (Figure 2F). These continuous subtype weights revealed that some individual maps were highly correlated with the subtypes, while others had only milder correlations, sometimes of similar amplitude for different subtypes. The subtype decomposition was therefore a discrete approximation of a continuous distribution of individual maps, rather than a set of clear-cut entities.

A comparison of clustering outcomes for the seven networks revealed that 3 subgroups of subjects at least could be evidenced in all networks (Figure 3A). As observed for the DMN, subtype maps showed distributed variations inside and outside the network of reference for all networks. While between-subject correlation values had similar amplitudes across networks, the size of the subgroups varied from one network to another. We tested the correspondence of subject clustering solutions between

networks by computing the adjusted rand index (ARI) for all pairwise comparisons (Figure 3B). The near-chance level of this metric (0.04 ± 0.04) demonstrated that subjects with similar connectivity maps for a given network did not have particularly similar maps for other networks, thus highlighting heterogeneity in functional brain connectivity patterns.

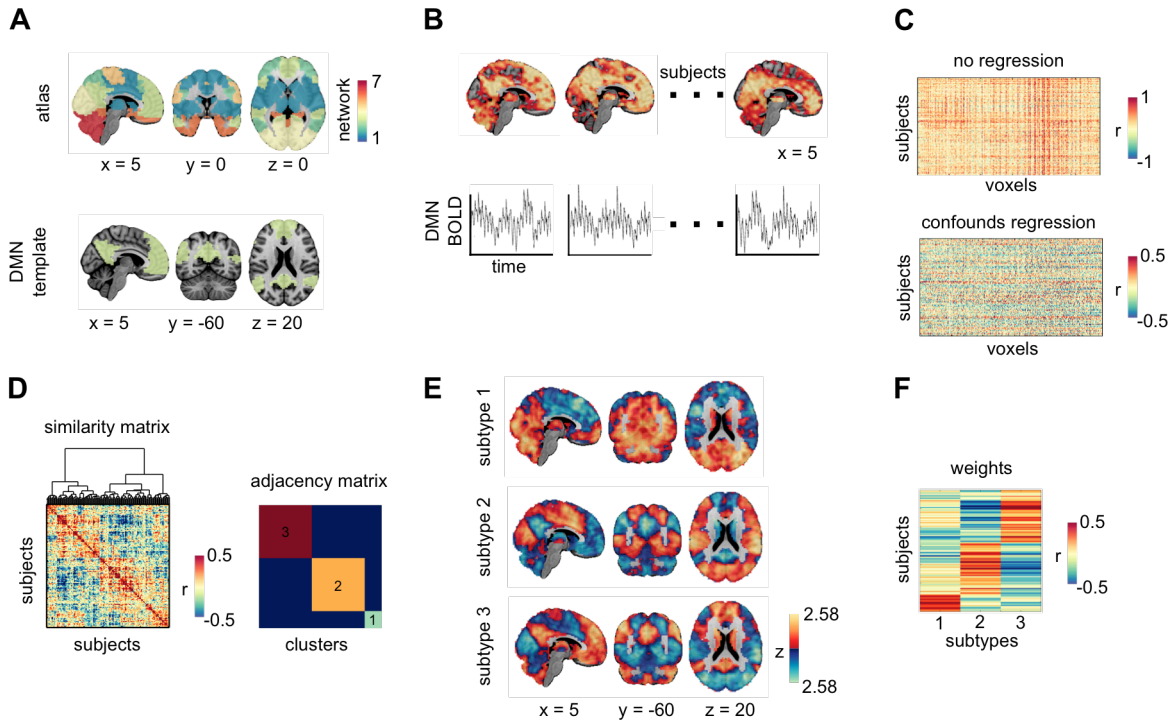

Figure 2. Extraction of subtypes and weights
(A) Functional subtypes were identified separately for 7 networks delineated at the whole-brain level in an independent sample of healthy subjects. The procedure is shown for the default-mode network (DMN). (B) Network-based connectivity maps were computed for each subject through the correlation of every voxel's time course of activity with the average signal in the reference network. (C) Site, gender, age and motion were regressed out from functional connectivity maps across subjects. (D) A hierarchical cluster analysis was conducted to identify 3 homogeneous subgroups of subjects with similar connectivity maps. (E) Difference subtypes show how the average connectivity maps of each separate subgroup of subjects differ from the grand average. (F) Weights consisted in correlations between the connectivity maps of every subject with that of each subtype.

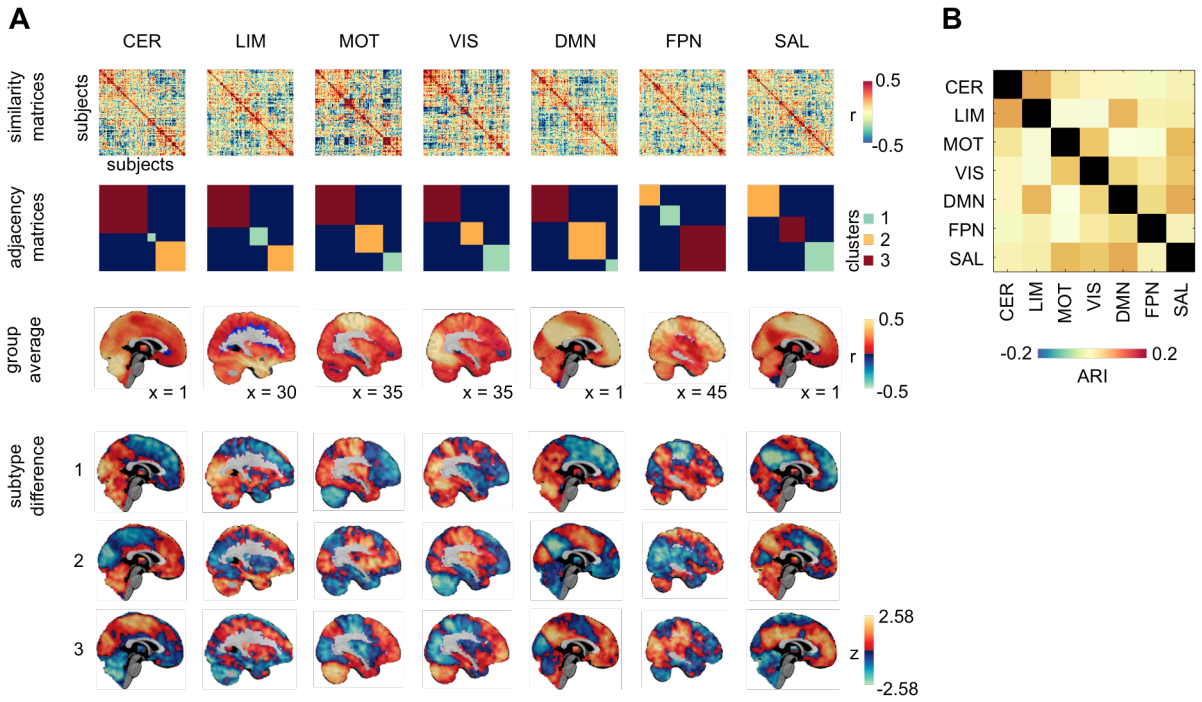

Figure 3. Correspondence of cluster (subtype) solutions across networks
(A) For each of the 7 networks (columns) are given the similarity matrix that shows the similarity of network connectivity maps between all pairs of subjects (first row), the adjacency matrix that reveals homogeneous subgroups of subjects identified by cluster analysis (second row), the average network connectivity map for all subjects (third row), and the difference subtype connectivity maps obtained by differences between the group average and the average connectivity maps for each subgroup of subjects (fourth to sixth rows). (B) The adjusted rand index (ARI) reveals the correspondence of subject clustering solutions between all pairs of networks. CER, cerebellum; LIM, limbic; MOT, motor; VIS, visual; DMN, default-mode; FPN, fronto-parietal; SAL, salience.

Brain network subtypes are associated with clinical symptoms

Given the observation that subtypes reflected both continuous and discrete phenomena, we adopted a dual statistical evaluation of their association with clinical symptoms in ADMCI subjects (Figure 4). In the former case, differences in average subtype weights between ADMCI and HC were assessed independently for each subtype of the seven reference networks, using a linear regression model. Significant associations were found for one limbic, two default-mode and two salience subtypes ($q < 0.05$ with FDR correction over 21 network subtypes), in line with our expectations. An uncorrected effect was also seen for an additional limbic subtype ($p < 0.05$). Effects were of medium size ($0.09 <$ Cohen's $f2 < 0.25$). Of these six subtypes, half of the associations with symptoms were positive (i.e. higher average weight load in ADMCI persons) and the remainder negative (i.e. lower average weight load in ADMCI patients). Instances of positive and negative associations with symptoms were observed in all three aforementioned networks.

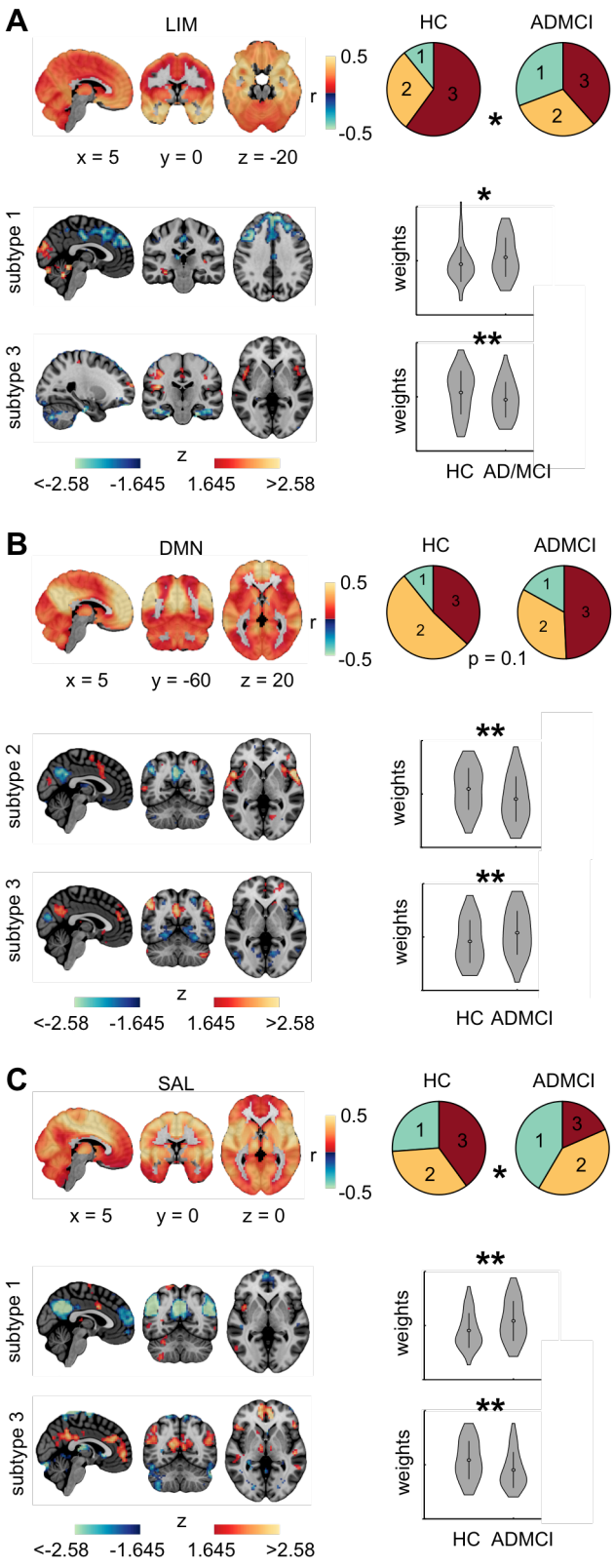

Figure 4. Functional network subtypes associated with clinical symptoms
Significant associations with ADMCI were found in the limbic (A), default-mode (B) and salience (C) networks. For each network are shown the group average connectivity map and the connectivity subtypes that are significantly more or less present in ADMCI patients than controls (difference maps are given). Pie charts report the distributions of subjects across subtypes in each group. Violin plots show the distribution of weights in the two groups for each subtype with a significant association. ** and * respectively denote significance at $q^{FDR}<0.05$ and $p<0.05$ (uncorrected).

A general observation was that subtypes positively associated with symptoms (PAS) had increased within-network connectivity but decreased between-network connectivity as compared to sample averages of networks. The PAS limbic subtype was notably defined by increased hippocampal connectivity (within-network) but decreased connectivity in dorsomedial prefrontal areas located in the DMN (between-network). An inverse pattern was seen in subtypes negatively associated with symptoms (NAS). The NAS limbic subtype had decreased connectivity in the hippocampus but increased connectivity in the insula. Subtypes of the default-mode and salience network provided mirror pictures of PAS and NAS connectivity profiles. Decreased connectivity in the posterior cingulate and medial prefrontal region relative to the sample average was NAS for the

default-mode network but PAS for the salience network. Similarly, decreased connectivity in the insula and anterior cingulate cortex compared to the sample average was evidenced to be NAS for the salience network but PAS for the default-mode network.

Statistics on discrete effects provided concordant effects at uncorrected thresholds. For each network, we evaluated with Chi2 tests whether ADMCI and HC subjects were distributed unevenly across subtypes. Unequal distributions were seen for the limbic ($p < 0.05$), default-mode ($p = 0.1$) and salience ($p < 0.05$) networks. Effect sizes were in the small-to-moderate range, with Cramer's V values of 0.27, 0.19 and 0.24 in the limbic, default-mode and salience networks, respectively.

Connectivity maps in FH subjects are reproducibly matched to subtypes from the clinical cohort

We assessed the reliability of matching connectivity maps in FH subjects from the PREVENT-AD cohort with the subtypes defined in the MTL-ADNI2. We thus generated individual functional connectivity maps separately for two runs, in each of the seven networks. Weights were computed for individual network maps, indicating their similarity with each of the 21 network subtypes previously defined in the MTL-ADNI2 sample (Figure 5). Intraclass correlations (ICC) indicated a fair-to-good correspondence of subtype weights between runs. Weights of all network subtypes had ICC values > 0.45 (max = 0.68, mean = 0.56), but for the PAS salience subtype (0.29). The default-mode and limbic PAS subtype weights had ICCs of 0.50 and 0.55, respectively.

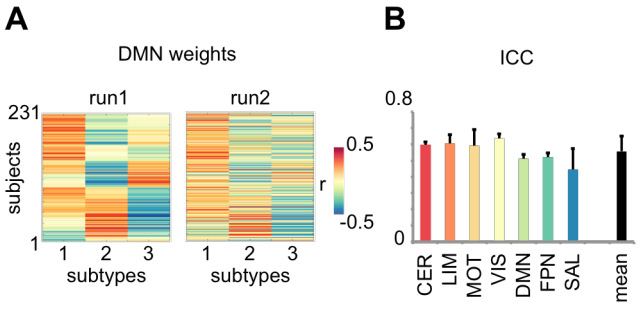

Figure 5. Reliability of subtype matching in FH subjects
(A) Matching of connectivity maps in FH subjects with subtypes found in the mixed population of ADMCI patients and controls is shown for the DMN in two separate runs. (B) Test-retest between runs was determined with intra-class correlation (ICC), showing fair-to-good correspondence across networks and subtypes.

Subtypes are associated with biomarkers of AD in FH subjects

We next examined the possibility that cognitively healthy FH elders already exhibited PAS subtypes, and more so than typical healthy elderly individuals. Individual functional connectivity maps were averaged for the two separate runs in 231 FH subjects from the PREVENT-AD cohort. For each of the three networks found to be associated with clinical symptoms, FH subjects were matched to network subtypes defined in the MTL-ADNI sample based on maximal weights. Distributions of FH subjects across subtypes

were not significantly different than those of either ADMCI or HC participants in the default-mode and salience networks (Figure 6A). However, proportions of FH subjects across limbic subtypes differed significantly from those of typical HC elders ($q < 0.05$) but not from ADMCI patients ($p = 0.9$).

The idea that connectivity subtypes might reflect a covert pathological AD process in cognitively healthy elderly individuals would be reinforced by the observation that such connectivity profiles correlate with known biomarkers of AD. We thus further investigated the relationship between connectivity subtypes and APOE genotype (N = 228) as well as CSF levels of $A\beta_{1\text{-}42}$, tTau and pTau (N = 59) (Figure 6C). Surprisingly, APOE allele 4 carriers showed less association than non carriers with the limbic PAS subtype ($q < 0.05$), with a small effect size (Cohen's f2 = 0.04). However, findings consistent with predictions were observed for CSF Aß42 levels and another limbic subtype. Subjects with high levels of CSF $A\beta_{1\text{-}42}$ had limbic connectivity maps that resembled more the NAS limbic network ($q < 0.05$; Cohen's f2 = 0.13). Low levels of CSF $A\beta_{1\text{-}42}$ were associated with another limbic subtype that shared some similarities with the at-risk limbic subtype, for instance increased hippocampal connectivity ($q < 0.05$; Cohen's f2 = 0.1). No associations were found between Tau or pTau CSF levels and any subtype of either the limbic, default-mode or salience networks.

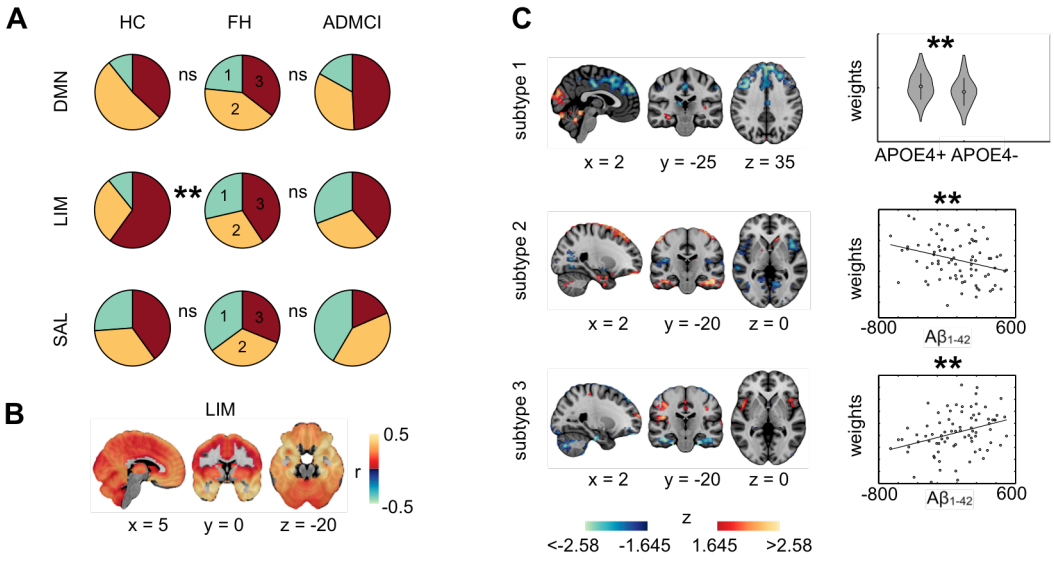

Figure 6. Connectivity subtypes in FH subjects
(A) Pie charts show that FH subjects differ from controls but not ADMCI patients in their distribution across subtypes for the limbic network (B). (C) Three distinct limbic network subtypes show either positive or negative associations with ApoE4 status or CSF $A\beta_{1\text{-}42}$ levels. ** denotes significance at $q^{FDR}<0.05$.

Discussion

Capturing heterogeneity through subtyping

Our subtyping approach was motivated by the lack of specificity and sensitivity of a clinical diagnosis of AD dementia against a histopathological diagnosis of AD pathology (Beach et al., 2012) and the variability of cognitive and neurobiological alterations in AD (Lam et al., 2013; Scheltens et al., 2016). As previously done for structural atrophy patterns (Dong et al., 2017; Hwang et al., 2016; Zhang et al., 2016) and white matter structural dysconnectivity (Doan et al., 2017), we employed a subtype analysis that identified subgroups of subjects sharing similar functional brain connectivity, in a fully data-driven way and irrespective of clinical diagnosis. This is an important conceptual difference with more traditional cross-sectional comparisons between clinical cohorts, which assumes some homogeneity in connectivity within each group, e.g. (Badhwar et al., 2017, 2016; Jones et al., 2016; Korolev et al., 2016). Improved characterization of the inherent heterogeneity of brain dysconnectivity in AD will ultimately facilitate more personalized diagnosis and treatment. This new line of inquiry is made possible by large neuroimaging databases such as the ADNI, and will become increasingly important with the emergence of populational cohorts with associated neuroimaging repositories, such as the UK biobank (Miller et al., 2016).

Association between connectivity subtypes and clinical symptoms

Using rs-fMRI, we identified functional brain connectivity subtypes associated both positively and negatively with symptoms. A variety of causal mechanisms may explain such associations, which may co-exist. An association may reflect the direct progression of AD neurodegeneration in the brain (Jones et al., 2016), the presence of comorbidities (Profenno et al., 2010), as well as some form of cognitive reserve, or lack thereof (Stern, 2006). The existence of an association in itself is not enough to disambiguate between these different interpretations. Associations between connectivity subtypes and symptoms were selectively detected in the default-mode, salience and limbic networks. These three networks have consistently been reported in the literature as altered in patients with AD dementia or MCI, see (Badhwar et al., 2017; Vemuri et al., 2012) for reviews. The associated subtype maps pointed at changes both within networks, e.g. higher intra-network connectivity in at-risk DMN subtype, and between networks, e.g. decreased inter-network connectivity in at-risk DMN subtypes with regions of the salience network.

Translation of connectivity subtypes from clinical to non-clinical individuals

The distribution of connectivity subtypes in a group of cognitively normal FH individuals was found to resemble more that seen in a patient group than control individuals. This observation was made only for the limbic network, but not the default-mode and salience networks. Assuming functional connectivity subtypes partly reflect the progression of AD pathology, finding early dysconnectivity in the limbic network is consistent with the Braak staging of neurodegeneration (Braak and Braak, 1991) and

the increased risk of sporadic AD due to family history (Tanzi, 2012). Conversely, the limbic subtype negatively associated with symptoms was under-represented in FH individuals, and was shown to positively associate with CSF $A\beta_{1-42}$ levels. Taken together, these associations support the notion that different subtypes of limbic connectivity reflect the progression of AD pathophysiology at a preclinical stage. A finding that was more difficult to interpret was that ApoE4 carriers had significantly less weight on the limbic subtype positively associated with symptoms. With previous literature on ApoE4 and resting-state connectivity reporting sometimes contradictory findings (Filippini et al., 2009; Sheline et al., 2010), we believe longitudinal data on a large cohort would be necessary to clarify the interplay of resting-state connectivity, Aβ deposition and ApoE4 status.

Generalization of brain connectivity subtypes across datasets

The translation of connectivity across cohorts raises the question of generalization across scanning sites. Research has indeed indicated that multisite scanning generates substantial site-specific bias in connectivity measures (Dansereau et al., 2017; Yan et al., 2013). In our multisite clinical sample, we took great care to control for confounding site effects on brain connectivity subtypes. The identification of network subtype was thus invariant to scanning site to a large extent. However, the cohort of individuals at risk of AD due to their familial history was entirely scanned at a distinct site. The fact that we found associations with known biomarkers or risk factors of AD specifically in the limbic network supports that brain connectivity subtypes are fairly robust to site effects. Subtype weights also had good test-retest reliability in the preventAD cohort, although the subtype maps were generated on ADNI-MTL. Important areas for future work will be to identify imaging protocols that further minimize differences in brain connectivity subtypes across scanners.

Finer subtypes

The subtypes found to positively associate with symptoms involved groups of patients which did not overlap a lot across networks. There is thus some degree of independence between subtypes, across networks, possibly reflecting heterogeneity of disease spread across patients. Even though we estimated only 3 subtypes per network, there are still a very large number of possible combinations of subtypes across 7 networks. Subtype maps being an average of a subgroup of subjects, a minimum number of 20 subjects seems warranted to stabilize the subtype maps. The total sample size of our discovery dataset thus constrained the maximal number of subtypes we could feasibly investigate. We thus decided to use low numbers of subtypes and networks for this first evaluation of the feasibility of functional subtypes in AD, yet higher numbers could be explored in a larger sample.

Multi-network and multimodal subtypes

A natural extension of this work would be to integrate subtypes across multiple networks, imaging modalities and measures into a single predictor of AD status.

Associations with clinical symptoms or AD biomarkers reported here had weak to moderate effect sizes, despite reaching statistical significance. Recent state-of-the-art model of progression from MCI to dementia indeed merge biomarkers across multiple domains, including cognitive evaluations, imaging and plasma markers (Korolev et al., 2016). High-dimensional imaging biomarkers such as structural and diffusion MRI are amenable to subtyping (Doan et al., 2017; Hwang et al., 2016; Zhang et al., 2016). We believe that subtyping could be used in the near future to identify a highly accurate multimodal predictor of AD, both for diagnosis and prognosis purposes. Resting-fMRI will likely contribute to such a multimodal predictor, as it is uniquely sensitive to brain function, at least compared to other MRI modalities. Our findings suggest that limbic subtypes in particular are promising biomarkers for the purpose of early AD diagnosis.

Conclusions

The present work demonstrates that rs-fMRI can be used to subtype the heterogeneity of functional networks in older adults. We found that subtypes have a good test-retest reliability and associate with symptoms in patients suffering from MCI or AD dementia. We also found that subtypes associate with various biomarkers and risk factors of AD in cognitively normal individuals: familial history of AD dementia, beta amyloid deposition, ApoE4 status. Our findings support the notion that rs-fMRI subtypes are sensitive to AD progression up to the preclinical stage, and may contribute to future efforts towards an accurate early diagnosis of AD using multimodal biomarkers.

Experimental procedures

Participants

The MTL-ADNI2 multisite sample aggregated data from 5 different studies: 3 samples from the Montreal area (one from the Montreal Neurological Institute, MNI, and two from the *Centre de Recherche de l'Institut Universitaire de Gériatrie de Montréal*, CRIUGMa and CRIUMGb), and 2 samples with distinct acquisition protocols from the Alzheimer's Disease Neuroimaging Initiative 2 (ADNI2a and ADNI2b) (Table 2). We selected subsamples of the MNI, CRIUGMa, CRIUGMb, ADNI2a and ADNI2b datasets such that patients and controls groups had identical sample size for each acquisition protocol or study, respectively 13, 13, 8, 20 and 11 subjects per group. The combined sample included 65 patients diagnosed with either amnestic MCI or AD dementia and 65 cognitively normal controls. Patients and controls were selected from a larger initial pool such that they would be matched for age, gender ratio as well as motion (see rs-fMRI preprocessing section). Distributions of age, gender and motion were as follows for patients vs. controls: age (mean ± std) = 72.7 ± 7.9 vs. 72.6 ± 7.3 years old, 41/24 vs. 41/24 females/males, residual frame displacement (mean ± std) = 0.22 ± 0.07 vs. 0.23 ± 0.08. All subjects gave informed consent to participate in these studies, which were approved by the research ethics committees of the institutions involved in data

acquisition. Consent was obtained for data sharing and secondary analysis, the later being approved by the ethics committee at the CRIUGM.

The PREVENT-AD dataset used in the present analysis included 231 cognitively healthy older adults with a known family history of AD, as reflected by a diagnosis of AD dementia in parent or multiple degree relatives. PREVENT-AD participants were younger (mean ± std: 64.1 ± 5.7 years old) than subjects in the MTL-ADNI2 multisite sample and were not balanced for gender (172/59 females/males). All subjects had given informed consent and the study was approved by the "Research, Ethics and Compliance Committee" of McGill University.

Note on the cohorts

The ADNI2 data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-profit organizations, as a $60 million, 5-year public-private partnership representing efforts of many co-investigators from a broad range of academic institutions and private corporations. A central goal of ADNI is to facilitate the discovery of biomarkers of very early AD progression, using MRI among other techniques. ADNI was followed by ADNI-GO and ADNI-2. In this study, we only included subjects from the two ADNI2 scanners (Achieva and Intera) associated with the largest samples. For up-to-date information, see www.adni-info.org.

The PREVENT-AD data were taken from the Pre-symptomatic Evaluation of Novel or Experimental Treatments for Alzheimer's Disease (PREVENT-AD) Cohort assembled at the Douglas Mental Health University Institute's Centre for Studies on Prevention of Alzheimer's Disease (StoP-AD) Centre, Montreal, Canada. This cohort was composed of cognitively healthy individuals at increased risk of AD dementia because they have / had a first-degree relative (parent or sibling) who has / had dementia suggestive of AD. This cohort includes volunteers of age 60 or older (55 or older if current age is within 15 years of affected relative's estimated age at onset of dementia). One current project consists in an observational study where participants are followed longitudinally once a year with a battery of tests and imaging modalities. In the present work, we focused on baseline data. A subset of test-retest rsfMRI data in 80 PREVENT-AD subjects has been shared publicly (Orban et al., 2015).

Clinical evaluation

All subjects from the MTL-ADNI2 and PREVENT-AD samples underwent neuropsychological testing to assess cognitive function, including memory, language and executive abilities. However, the neuropsychological tests administered to participants varied across sites, as did criteria and clinical scales used for diagnosis of either MCI or AD. Briefly, patients with (amnestic) MCI had memory complaints and objective cognitive loss, yet showed intact functional abilities and did not meet criteria

for a diagnosis of dementia in contrast with AD patients. HC demonstrated intact cognitive functions. Details on clinical evaluation for each cohort per site follow.

In ADNI2, the Mini-Mental State Evaluation (MMSE) and Clinical Dementia Rating (CDR) were used to distinguish between HC, MCI and AD subjects. MMSE scores were inclusively comprised between 24-30, 24-30 and 20-26 for HC, MCI and AD subjects, respectively. MCI patients had a CDR of 0.5 and AD patients a CDR of 0.5 or 1. An objective memory loss was evidenced with the Wechsler Memory Scale Logical Memory II in MCI, yet other cognitive domains and functional activities were unaffected. In addition, there was an absence of dementia, by contrast with AD patients who met the National Institute of Neurological and Communicative Disorders and Stroke / Alzheimer's Disease and Related Disorders Association (NINCDS/ADRDA) criteria for probable AD (McKhann et al., 1984). The MNI sample only included MCI patients, who were similarly diagnosed using the MMSE, following Petersen Criteria (Petersen, 2004). Subjects in the CRIUGM samples were administered the MMSE as well as the Montreal Cognitive Assessment (MoCA) (Nasreddine et al., 2005) and the Mattis Dementia Rating Scale (MDRS, Mattis et al., 1988). The diagnosis of MCI was made based on scores equal to or >1.5 standard deviations below the mean adjusted for age and education on memory tests, with input from a neurologist. A diagnosis of AD was determined according to the Diagnostic and Statistical Manual of Mental Disorders (4th ed.; American Psychiatric Association, 2000) and NINCDS/ADRDA clinical criteria, with input from a neurologist. Participants in the PREVENT-AD were evaluated for any cognitive impairment and symptoms suggestive of AD using the Repeatable Battery for the Assessment of Neuropsychological Status - RBANS (Randolph et al., 1998), the CDR, the MoCA and the AD8 Dementia screening (Galvin et al., 2005). Exclusion criteria common to all participants included contraindications to MRI, presence or history of axis I psychiatric disorders (e.g., depression, bipolar disorder, schizophrenia), presence or history of neurologic disease with potential impact on cognition (e.g., Parkinson’s disease), and presence or history of substance abuse.

Genetic and CSF biomarkers in PREVENT-AD subjects

In 228 PREVENT-AD subjects, DNA was isolated from 200 ul of whole blood using a QIASymphony apparatus and the DNA Blood Mini QIA Kit (Qiagen, Valencia, CA, USA). The standard QIASymphony isolation program was performed as per the manufacturer's instructions. APOE single nucleotide polymorphism (SNP) genotyping was performed using pyrosequencing (PyroMArk96) and processed with GenomeStudio (version 2010.3) using standard methods .

CSF samples were obtained by lumbar puncture in 59 subjects of the PREVENT-AD cohort. For each subject, 25 ml of CSF was centrifuged 10 minutes +/- 2000g at room temperature and aliquoted in 50 vials of 0.5 ml and frozen at -80C for further analysis. Protein levels of $A\beta_{1-42}$, total tau (tTau) and phosphorylated tau (pTau) were determined by enzyme-linked immunosorbent assay (ELISA) from Innotest technology (Fujirebio). These measurements were standardized with the European project

BIOMARKAPD (Reijs et al., 2015), which intends to harmonize assays that are used to measure biological markers in neurodegenerative diseases.

MRI acquisition

The MTL-ADNI2 multisite resting-state dataset included brain imaging data acquired on 3T MRI scanners. Vendors differed between sites (Siemens Magnetom Tim Trio in MTL sites and Phillips Achieva or Intera in ADNI2). Analyses were performed on the first usable scan, typically the baseline scan when several scans were available. Functional scan acquisition parameters varied from one site to another, notably in run duration (ranges: 5min20s-8min), number of volume (range: 140-240 vols), voxel size (range: 3-4x3-3.6x3.3-4mm^3) and repetition time (range: 2-3s). Brain imaging data of the PREVENT-AD dataset were collected on a single 3T MRI scanner (Siemens, Magnetom Tim Trio). Two consecutive resting-state runs of 150 functional volumes were acquired, each run lasting 5min 45s. Spatial and temporal resolutions were as follows: voxel size = 4x4x4mm^3 and repetition time = 2000ms. Table 2 reports scan acquisition parameters for all data.

rs-fMRI preprocessing

Datasets were preprocessed and analyzed using the NeuroImaging Analysis Kit - NIAK - version 0.12.17 (http://niak.simexp-lab.org), under CentOS with Octave (http://gnu.octave.org) version 3.6.1 and the MINC toolkit (http://bic-mni.github.io/) version 0.3.18. Analyses were executed in parallel on the "Guillimin" supercomputer (http://www.calculquebec.ca/en/resources/compute-servers/guillimin), using the pipeline system for Octave and Matlab - PSOM (Bellec et al., 2012).

Each fMRI dataset 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., 2011), 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 frames showing a displacement greater than 0.5 mm were removed (Power et al., 2012). The following nuisance covariates were regressed out from the fMRI 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; Lund et al., 2006). The fMRI volumes were finally spatially smoothed with a 6 mm isotropic Gaussian blurring kernel. A more detailed description of the pipeline can be found on the NIAK website (http://niak.simexp-lab.org/pipe_preprocessing.html).

Individual voxel-wise connectivity maps based on large-scale network templates

For all 361 subjects included in the analyses, we computed voxel-wise connectivity maps associated with each of 7 network templates extracted from a functional brain atlas generated on 200 healthy subjects (https://doi.org/10.6084/m9.figshare.1285615.v1). The atlas included cerebellar, limbic, visual, motor, default-mode, fronto-parietal and salience networks. For each subject and each network, a network connectivity map was obtained by computing the Fisher-transformed Pearson's correlations between the average time course within the network template and the time course of every voxel in the brain grey matter. For each network, subject by voxel connectivity matrices were defined at the group level, separately for the MTL-ADNI and PREVENT-AD samples. Two general linear models were used to regress the following confounds on the group connectivity matrices: age, sex and rFD, as well as acquisition protocols / study using dummy variables, i.e. MNI, CRIUGMa, CRIUGMb, ADNIa, ADNIb. The inclusion of constant terms in the models effectively normalized network connectivity maps to a zero grand mean across all subjects, separately for the MTL-ADNI and PREVENT-AD samples.

Network subtypes defined by a cluster analysis in MTL-ADNI2 subjects

For each network, a subject by subject similarity matrix summarized the between-subject correspondence of connectivity maps for all pairs of the 130 subjects in the MTL-ADNI multisite sample. A hierarchical cluster analysis was performed to identify 3 clusters of subjects whose network connectivity maps were similar in terms of spatial extent and/or strength. For each cluster, we defined a subtype of functional connectivity as the average connectivity map for subjects within this cluster. Subtype weights were obtained by calculating the correlation between individual connectivity maps and each of the network subtype maps. Weights thus range between -1 and 1, with 1 meaning perfect correspondence, 0 lack of correspondence and -1 perfect but inverted correspondence.

Statistical tests of association with clinical symptoms in MTL-ADNI2 subjects

We tested the association between subtypes of network connectivity and clinical symptoms in the 130 MTL-ADNI2 subjects. To this end, we employed two distinct statistical approaches: one approach treated subtypes as discrete units, where each subject belongs to one and only one cluster; a second approach used subtype weights, which are continuous measures. Despite these conceptual differences, we expected both statistical approaches to provide mostly concordant results. In the first approach, Chi2 tests were used to reveal unequal distributions of HC and ADMCI patients across the subtype clusters of each network. We report Cramer's V effect sizes for which values of 0.1, 0.3 and 0.5 are respectively termed small, medium and large. In our second approach, we used a general linear models to test separately the associations between the weights of each network subtype and clinical symptoms (HC vs. ADMCI). Because confounds (age, sex, rFD, sites) were regressed out prior to conduct this

analysis, no factors of interest were entered in the general linear model. We provide Cohen's f2 effect sizes for which values of 0.02, 0.15 and 0.35 are termed small, medium and large, respectively (Cohen, 1988). In both statistical approaches, results were deemed significant if they survived false-discovery rate (FDR) correction at q<0.05 across networks and subtypes.

Matching of FH subjects to at-risk subtypes

We next aimed to match connectivity maps in 231 cognitively normal FH elders with at-risk subtypes identified in the MTL-ADNI2 dataset. For each network and each PREVENT-AD subject, subtype weights were obtained by correlating his/her connectivity map (averaged over 2 runs) with each of the 3 subtype maps identified in the clinical sample. Each FH subject was assigned to the subtype for which the weight was maximal. We then tested, for each network, the similarity of subject distributions across subtypes between FH subjects in the PREVENT-AD cohort vs the distribution of ADMCI patients or HC subjects in the MTL-ADNI multisite sample. Chi2 tests were used to assess significance of differences in distributions and Cramer's V values described effect sizes.

Test-retest reliability of MTL-ADNI2 subtypes in FH subjects

Intra-class correlation coefficients quantified the reproducibility of weights between the two consecutive resting-state runs of the PREVENT-AD cohort. With 7 networks and 3 subtypes, we thus obtained 21 ICC measures. ICC measures were interpreted as follows (Cicchetti, 1994): less than 0.40 = poor, between 0.40 and 0.59 = fair, between .60 and 0.74 = good, between 0.75 and 1 = excellent.

Statistical tests of association with AD biomarkers

We finally assessed whether the subtype weights of FH subjects would be associated with known biomarkers or risk factors of AD in PREVENT-AD. Namely, we investigated the possible association between APOE4 genotype, CSF $A\beta_{1-42}$ and Tau levels as well as spEYO with either at-risk or protective network connectivity subtypes. Associations were tested in the framework of general linear models and were considered significant if they survived false-discovery rate (FDR) correction at q<0.05 across networks and subtypes. Because confounds (age, sex, rFD) were regressed out prior to conduct this analysis, no factors of interest were entered in the general linear models. Effect sizes are reported with Cohen's f2 measures.

Supplemental Information

NA

Author contributions

(...)

Acknowledgments

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Alzheimer's Association; Alzheimer's Drug Discovery Foundation; BioClinica, Inc.; Biogen Idec Inc.; Bristol-Myers Squibb Company; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; GE Healthcare; Innogenetics, N.V.; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Medpace, Inc.; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Synarc Inc.; and Takeda Pharmaceutical Company. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California. This research was also supported by NIH grants P30 AG010129 and K01 AG030514.

This work was supported by a salary award (Junior 1 scholarship) by Fonds de Recherche du Québec - Santé (FRQS) as well as funds by the Canadian Institutes of Health research (CIHR), the Alzheimer’s Society of Canada and the Courtois foundation to PB. AB was supported by postdoctoral fellowships from the Canadian Alzheimer Society and CIHR, as well as by the Lemaire foundation. CD is supported by a bursary from the Lemaire foundation. The PREVENT-AD cohort was funded by generous support from McGill University, the government of Canada, an unrestricted gift from Pfizer Canada, the Canada Fund for Innovation, the Douglas Hospital Research Centre,

the Levesque Foundation, McGill University and Genome Quebec Innovation Center. We are grateful to Sylvie Belleville, Howard Chertkow, Louis Collins, Samir Das, Alain Dagher, Alan Evans, Alexandru Hanganu, Ouri Monchi, Amir Schmuel, Louise Theroux and Sequian Wang for sharing fMRI datasets and/or analytical tools. Finally, we would like to acknowledge the participants of the Prevent-AD cohort for dedicating their time and energy to help us collecting this data.

Tables

		MTL-ADNI2				PREVENT-AD
	MNI	**CRIUGMa**	**CRIUGMb**	**ADNI2a**	**ADNI2b**	
N controls	13	13	8	20	11	n/a
Mean age (s.d.)	67 (5.8)	71.2 (4.8)	72.6 (7.8)	75.3 (6.5)	75.9 (8.7)	n/a
Number male (%)	5 (38.5)	4 (30.8)	5 (62.5)	9 (45)	1 (9.1)	n/a
N ADMCI patients	13	13	8	20	11	n/a
N MCI patients	13	0	8	13	10	n/a
N AD dementia patients	0	13	0	7	1	n/a
Mean age (s.d.)	71.6 (8.4)	75 (7)	79.9 (6.1)	72 (7.9)	67 (5)	n/a
Number male (%)	5 (38.5)	2 (15.4)	3 (37.5)	7 (35)	7 (63.6)	n/a
N FH subjects	n/a	n/a	n/a	n/a	n/a	231
Mean age (s.d.)	n/a	n/a	n/a	n/a	n/a	64.1 (5.7)
Number male (%)	n/a	n/a	n/a	n/a	n/a	59 (25.5)
N $A\beta_{1-42}$	n/a	n/a	n/a	n/a	n/a	79
Mean $A\beta_{1-42}$ (s.d.)	n/a	n/a	n/a	n/a	n/a	1079.7 (280.9)
N ApoE4	n/a	n/a	n/a	n/a	n/a	228
N ApoE4 carriers (%)	n/a	n/a	n/a	n/a	n/a	78 (34.2)

Table 1. Demographics

Basic demographics (sample size, mean age, sex proportions) are given for the HC, ADMCI and FH groups. Levels of CSF $A\beta_{1-42}$ and proportions of ApoE4 carriers are given for FH subjects.

	MTL-ADNI2					PREVENT-AD
	MNI	**CRIUGMa**	**CRIUGMb**	**ADNI2a**	**ADNI2b**	
Scanner manufacturer	Siemens	Siemens	Siemens	Phillips	Phillips	Siemens
Structural						
N channels	32	32	32	8	8	12
N slices	176	176	176	170	170	176
Voxel size (mm3)	1x1x1	1x1x1	1x1x1	1x1x1.2	1x1x1.2	1x1x1
Matrix size	256x256	256x256	240x256	256x256	256x256	256x256
FOV (mm2)	256	256	240/256	256	256	256?
TR (s)	2.3	2.53	2.3	6.8	6.8	2.3
TE (ms)	2.98	1.64	2.91	3.09	3.09	2.98
FA (degrees)	9	9	9	9	9	9
Functional						
N channels	32	32	32	8	8	12
N slices	38	33	33	48	48	32
Voxel size (mm3)	3.6x3.6x3.6	3x3x4	3x3x4	3.3x3.3x3.3	3.3x3.3x3.3	4x4x4
Matrix size	64x64	64x64	64x64	64x64	64x64	64x64
FOV (mm2)	230	192	192	212	212	256?
TR (s)	2	2	2	3	3	2
TE (ms)	30	30	30	30	30	30
FA (degrees)	90	90	90	80	80	90
No. volumes	160	240	240	140	140	150 (x 2)
Scan duration (min:s)	5:20	8:00	8:00	7:00	7:00	5:45 (x2)

Table 2. MRI acquisition protocols

Scan parameters are given for structural and functional data across the 5 MTL-ADNI samples as well as the PREVENT-AD dataset.

References

Ad-Dab'bagh, Y., Lyttelton, O., Muehlboeck, J.S., Lepage, C., Einarson, D., Mok, K., Ivanov, O., Vincent, R.D., Lerch, J., Fombonne, E., Others, 2006. 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. Florence, Italy, p. 2266.

Badhwar, A., Tam, A., Dansereau, C., Orban, P., Hoffstaedter, F., Bellec, P., 2017. Resting-state network dysfunction in Alzheimer's disease: A systematic review and meta-analysis. Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring 8, 73–85.

Badhwar, A., Tam, A., Dansereau, C., Orban, P., Toro, R., Bellec, P., 2016. Resting-state network dysfunction in Alzheimer's disease: a systematic review and meta-analysis. Alzheimers. Dement.

Beach, T.G., Monsell, S.E., Phillips, L.E., Kukull, W., 2012. Accuracy of the clinical diagnosis of Alzheimer disease at National Institute on Aging Alzheimer Disease Centers, 2005-2010. J. Neuropathol. Exp. Neurol. 71, 266–273.

Bellec, P., Lavoie-Courchesne, S., Dickinson, P., Lerch, J.P., Zijdenbos, A.P., Evans, A.C., 2012. The pipeline system for Octave and Matlab (PSOM): a lightweight scripting framework and execution engine for scientific workflows. Front. Neuroinform. 6, 7.

Braak, H., Braak, E., 1991. Neuropathological stageing of Alzheimer-related changes. Acta Neuropathol. 82, 239–259.

Brier, M.R., Thomas, J.B., Ances, B.M., 2014. Network dysfunction in Alzheimer's disease: refining the disconnection hypothesis. Brain Connect. 4, 299–311.

Buckner, R.L., Snyder, A.Z., Shannon, B.J., LaRossa, G., Sachs, R., Fotenos, A.F., Sheline, Y.I., Klunk, W.E., Mathis, C.A., Morris, J.C., Mintun, M.A., 2005. Molecular, structural, and functional characterization of Alzheimer's disease: evidence for a relationship between default activity, amyloid, and memory. J. Neurosci. 25, 7709–7717.

Chételat, G., La Joie, R., Villain, N., Perrotin, A., de La Sayette, V., Eustache, F., Vandenberghe, R., 2013. Amyloid imaging in cognitively normal individuals, at-risk populations and preclinical Alzheimer's disease. Neuroimage Clin 2, 356–365.

Cicchetti, D.V., 1994. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychol. Assess. 6, 284.

Cohen, J., 1988. Statistical power analysis for the behavioral sciences Lawrence Earlbaum Associates. Hillsdale, NJ 20–26.

Collins, D.L., Evans, A.C., 1997. Animal: Validation and Applications of Nonlinear Registration-Based Segmentation. Int. J. Pattern Recognit Artif Intell. 11, 1271–1294.

Dansereau, C., Benhajali, Y., Risterucci, C., Pich, E.M., Orban, P., Arnold, D., Bellec, P., 2017. Statistical power and prediction accuracy in multisite resting-state fMRI connectivity. Neuroimage 149, 220–232.

Delbeuck, X., Van der Linden, M., Collette, F., 2003. Alzheimer'disease as a disconnection syndrome? Neuropsychol. Rev. 13, 79–92.

Doan, N.T., Engvig, A., Persson, K., Alnæs, D., Kaufmann, T., Rokicki, J., Córdova-

Palomera, A., Moberget, T., Brækhus, A., Barca, M.L., Engedal, K., Andreassen, O.A., Selbæk, G., Westlye, L.T., 2017. Dissociable diffusion MRI patterns of white matter microstructure and connectivity in Alzheimer's disease spectrum. Sci. Rep. 7, 45131.

Dong, A., Toledo, J.B., Honnorat, N., Doshi, J., Varol, E., Sotiras, A., Wolk, D., Trojanowski, J.Q., Davatzikos, C., Alzheimer's Disease Neuroimaging Initiative, 2017. Heterogeneity of neuroanatomical patterns in prodromal Alzheimer's disease: links to cognition, progression and biomarkers. Brain 140, 735–747.

Dubois, B., Hampel, H., Feldman, H.H., Scheltens, P., Aisen, P., Andrieu, S., Bakardjian, H., Benali, H., Bertram, L., Blennow, K., Broich, K., Cavedo, E., Crutch, S., Dartigues, J.-F., Duyckaerts, C., Epelbaum, S., Frisoni, G.B., Gauthier, S., Genthon, R., Gouw, A.A., Habert, M.-O., Holtzman, D.M., Kivipelto, M., Lista, S., Molinuevo, J.-L., O'Bryant, S.E., Rabinovici, G.D., Rowe, C., Salloway, S., Schneider, L.S., Sperling, R., Teichmann, M., Carrillo, M.C., Cummings, J., Jack, C.R., Jr, Proceedings of the Meeting of the International Working Group (IWG) and the American Alzheimer's Association on "The Preclinical State of AD"; July 23, 2015; Washington DC, USA, 2016. Preclinical Alzheimer's disease: Definition, natural history, and diagnostic criteria. Alzheimers. Dement. 12, 292–323.

Elman, J.A., Madison, C.M., Baker, S.L., Vogel, J.W., Marks, S.M., Crowley, S., O'Neil, J.P., Jagust, W.J., 2016. Effects of Beta-Amyloid on Resting State Functional Connectivity Within and Between Networks Reflect Known Patterns of Regional Vulnerability. Cereb. Cortex 26, 695–707.

Filippini, N., MacIntosh, B.J., Hough, M.G., Goodwin, G.M., Frisoni, G.B., Smith, S.M., Matthews, P.M., Beckmann, C.F., Mackay, C.E., 2009. Distinct patterns of brain activity in young carriers of the APOE-ε4 allele. Proceedings of the National Academy of Sciences 106, 7209–7214.

Fonov, V., Evans, A.C., Botteron, K., Almli, C.R., McKinstry, R.C., Collins, D.L., Brain Development Cooperative Group, 2011. Unbiased average age-appropriate atlases for pediatric studies. Neuroimage 54, 313–327.

Galvin, J.E., Roe, C.M., Powlishta, K.K., Coats, M.A., Muich, S.J., Grant, E., Miller, J.P., Storandt, M., Morris, J.C., 2005. The AD8: a brief informant interview to detect dementia. Neurology 65, 559–564.

Giove, F., Gili, T., Iacovella, V., Macaluso, E., Maraviglia, B., 2009. Images-based suppression of unwanted global signals in resting-state functional connectivity studies. Magn. Reson. Imaging 27, 1058–1064.

Greicius, M.D., Srivastava, G., Reiss, A.L., Menon, V., 2004. Default-mode network activity distinguishes Alzheimer's disease from healthy aging: evidence from functional MRI. Proc. Natl. Acad. Sci. U. S. A. 101, 4637–4642.

Hwang, J., Kim, C.M., Jeon, S., Lee, J.M., Hong, Y.J., Roh, J.H., Lee, J.-H., Koh, J.-Y., Na, D.L., Alzheimer's Disease Neuroimaging Initiative, 2016. Prediction of Alzheimer's disease pathophysiology based on cortical thickness patterns. Alzheimers. Dement. 2, 58–67.

Hyman, B.T., Phelps, C.H., Beach, T.G., Bigio, E.H., Cairns, N.J., Carrillo, M.C., Dickson, D.W., Duyckaerts, C., Frosch, M.P., Masliah, E., Mirra, S.S., Nelson, P.T., Schneider, J.A., Thal, D.R., Thies, B., Trojanowski, J.Q., Vinters, H.V., Montine, T.J., 2012. National Institute on Aging-Alzheimer's Association guidelines for the

neuropathologic assessment of Alzheimer's disease. Alzheimers. Dement. 8, 1–13.
Jiang, Y., Huang, H., Abner, E., Broster, L.S., Jicha, G.A., Schmitt, F.A., Kryscio, R., Andersen, A., Powell, D., Van Eldik, L., Gold, B.T., Nelson, P.T., Smith, C., Ding, M., 2016. Alzheimer's Biomarkers are Correlated with Brain Connectivity in Older Adults Differentially during Resting and Task States. Front. Aging Neurosci. 8, 15.
Jones, D.T., Knopman, D.S., Gunter, J.L., Graff-Radford, J., Vemuri, P., Boeve, B.F., Petersen, R.C., Weiner, M.W., Jack, C.R., Jr, Alzheimer's Disease Neuroimaging Initiative, 2016. Cascading network failure across the Alzheimer's disease spectrum. Brain 139, 547–562.
Korolev, I.O., Symonds, L.L., Bozoki, A.C., Alzheimer's Disease Neuroimaging Initiative, 2016. Predicting Progression from Mild Cognitive Impairment to Alzheimer's Dementia Using Clinical, MRI, and Plasma Biomarkers via Probabilistic Pattern Classification. PLoS One 11, e0138866.
Lam, B., Masellis, M., Freedman, M., Stuss, D.T., Black, S.E., 2013. Clinical, imaging, and pathological heterogeneity of the Alzheimer's disease syndrome. Alzheimers. Res. Ther. 5, 1.
Lund, T.E., Madsen, K.H., Sidaros, K., Luo, W.-L., Nichols, T.E., 2006. Non-white noise in fMRI: does modelling have an impact? Neuroimage 29, 54–66.
Miller, K.L., Alfaro-Almagro, F., Bangerter, N.K., Thomas, D.L., Yacoub, E., Xu, J., Bartsch, A.J., Jbabdi, S., Sotiropoulos, S.N., Andersson, J.L.R., Griffanti, L., Douaud, G., Okell, T.W., Weale, P., Dragonu, I., Garratt, S., Hudson, S., Collins, R., Jenkinson, M., Matthews, P.M., Smith, S.M., 2016. Multimodal population brain imaging in the UK Biobank prospective epidemiological study. Nat. Neurosci. 19, 1523–1536.
Mufson, E.J., Malek-Ahmadi, M., Perez, S.E., Chen, K., 2016. Braak staging, plaque pathology, and APOE status in elderly persons without cognitive impairment. Neurobiol. Aging 37, 147–153.
Nasreddine, Z.S., Phillips, N.A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J.L., Chertkow, H., 2005. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. J. Am. Geriatr. Soc. 53, 695–699.
Power, J.D., Barnes, K.A., Snyder, A.Z., Schlaggar, B.L., Petersen, S.E., 2012. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. Neuroimage 59, 2142–2154.
Profenno, L.A., Porsteinsson, A.P., Faraone, S.V., 2010. Meta-analysis of Alzheimer's disease risk with obesity, diabetes, and related disorders. Biol. Psychiatry 67, 505–512.
Randolph, C., Tierney, M.C., Mohr, E., Chase, T.N., 1998. The Repeatable Battery for the Assessment of Neuropsychological Status (RBANS): preliminary clinical validity. J. Clin. Exp. Neuropsychol. 20, 310–319.
Reijs, B.L.R., Teunissen, C.E., Goncharenko, N., Betsou, F., Blennow, K., Baldeiras, I., Brosseron, F., Cavedo, E., Fladby, T., Froelich, L., Gabryelewicz, T., Gurvit, H., Kapaki, E., Koson, P., Kulic, L., Lehmann, S., Lewczuk, P., Lleó, A., Maetzler, W., de Mendonça, A., Miller, A.-M., Molinuevo, J.L., Mollenhauer, B., Parnetti, L., Rot, U., Schneider, A., Simonsen, A.H., Tagliavini, F., Tsolaki, M., Verbeek, M.M., Verhey, F.R.J., Zboch, M., Winblad, B., Scheltens, P., Zetterberg, H., Visser, P.J.,

2015. The Central Biobank and Virtual Biobank of BIOMARKAPD: A Resource for Studies on Neurodegenerative Diseases. Front. Neurol. 6, 216.
Scheltens, N.M.E., Galindo-Garre, F., Pijnenburg, Y.A.L., van der Vlies, A.E., Smits, L.L., Koene, T., Teunissen, C.E., Barkhof, F., Wattjes, M.P., Scheltens, P., van der Flier, W.M., 2016. The identification of cognitive subtypes in Alzheimer's disease dementia using latent class analysis. J. Neurol. Neurosurg. Psychiatry 87, 235–243.
Seeley, W.W., Crawford, R.K., Zhou, J., Miller, B.L., Greicius, M.D., 2009. Neurodegenerative diseases target large-scale human brain networks. Neuron 62, 42–52.
Selkoe, D.J., 2002. Alzheimer's disease is a synaptic failure. Science 298, 789–791.
Sheline, Y.I., Morris, J.C., Snyder, A.Z., Price, J.L., Yan, Z., D'Angelo, G., Liu, C., Dixit, S., Benzinger, T., Fagan, A., Goate, A., Mintun, M.A., 2010. APOE4 allele disrupts resting state fMRI connectivity in the absence of amyloid plaques or decreased CSF Aβ42. J. Neurosci. 30, 17035–17040.
Sperling, R.A., Karlawish, J., Johnson, K.A., 2012. Preclinical Alzheimer disease—the challenges ahead. Nat. Rev. Neurol. 9, 54–58.
Stern, Y., 2006. Cognitive reserve and Alzheimer disease. Alzheimer Dis. Assoc. Disord. 20, S69–74.
Tampellini, D., 2015. Synaptic activity and Alzheimer's disease: a critical update. Front. Neurosci. 9, 423.
Tanzi, R.E., 2012. The genetics of Alzheimer disease. Cold Spring Harb. Perspect. Med. 2.
Vemuri, P., Jones, D.T., Jack, C.R., Jr, 2012. Resting state functional MRI in Alzheimer's Disease. Alzheimers. Res. Ther. 4, 2.
Wang, L., Brier, M.R., Snyder, A.Z., Thomas, J.B., Fagan, A.M., Xiong, C., Benzinger, T.L., Holtzman, D.M., Morris, J.C., Ances, B.M., 2013. Cerebrospinal fluid Aβ42, phosphorylated Tau181, and resting-state functional connectivity. JAMA Neurol. 70, 1242–1248.
Wang, L., Roe, C.M., Snyder, A.Z., Brier, M.R., Thomas, J.B., Xiong, C., Benzinger, T.L., Morris, J.C., Ances, B.M., 2012. Alzheimer disease family history impacts resting state functional connectivity. Ann. Neurol. 72, 571–577.
Yan, C.-G., Craddock, R.C., Zuo, X.-N., Zang, Y.-F., Milham, M.P., 2013. Standardizing the intrinsic brain: towards robust measurement of inter-individual variation in 1000 functional connectomes. Neuroimage 80, 246–262.
Zhang, X., Mormino, E.C., Sun, N., Sperling, R.A., Sabuncu, M.R., Yeo, B.T.T., Weiner, M.W., Aisen, P., Weiner, M., Petersen, R., Others, 2016. Bayesian model reveals latent atrophy factors with dissociable cognitive trajectories in Alzheimer's disease. Proceedings of the National Academy of Sciences 113, E6535–E6544.