



UNIL | Université de Lausanne

Unicentre

CH-1015 Lausanne

<http://serval.unil.ch>

Year : 2017

APPLICATION OF NEUROIMAGING AND GENETICS ON ALZHEIMER'S DISEASE

Cui Jing

Cui Jing, 2017, APPLICATION OF NEUROIMAGING AND GENETICS ON ALZHEIMER'S DISEASE

Originally published at : Thesis, University of Lausanne

Posted at the University of Lausanne Open Archive <http://serval.unil.ch>

Document URN : urn:nbn:ch:serval-BIB_6D40382D865D0

Droits d'auteur

L'Université de Lausanne attire expressément l'attention des utilisateurs sur le fait que tous les documents publiés dans l'Archive SERVAL sont protégés par le droit d'auteur, conformément à la loi fédérale sur le droit d'auteur et les droits voisins (LDA). A ce titre, il est indispensable d'obtenir le consentement préalable de l'auteur et/ou de l'éditeur avant toute utilisation d'une oeuvre ou d'une partie d'une oeuvre ne relevant pas d'une utilisation à des fins personnelles au sens de la LDA (art. 19, al. 1 lettre a). A défaut, tout contrevenant s'expose aux sanctions prévues par cette loi. Nous déclinons toute responsabilité en la matière.

Copyright

The University of Lausanne expressly draws the attention of users to the fact that all documents published in the SERVAL Archive are protected by copyright in accordance with federal law on copyright and similar rights (LDA). Accordingly it is indispensable to obtain prior consent from the author and/or publisher before any use of a work or part of a work for purposes other than personal use within the meaning of LDA (art. 19, para. 1 letter a). Failure to do so will expose offenders to the sanctions laid down by this law. We accept no liability in this respect.



UNIL | Université de Lausanne

Faculté de biologie
et de médecine

Département de Neurosciences Cliniques

**APPLICATION OF NEUROIMAGING AND GENETICS ON
ALZHEIMER'S DISEASE**

Thèse de doctorat en Neurosciences

présentée à la

Faculté de Biologie et de Médecine
de l'Université de Lausanne

par

Jing Cui

Bio-informatique diplômée de l'Université de la Sarre, Allemagne

Jury

Prof. Jean-Pierre Hornung, Président

Dr. Ferath Kherif, Directeur

Prof. Zoltán Kutalik, Co-Directeur

Prof. Giovanni Frisoni, Expert

Mr. Jérôme Dauvillier, Expert

Thèse n° 195

Lausanne 2016

*Programme doctoral interuniversitaire en Neurosciences
des Universités de Lausanne et Genève*



ACKNOWLEDGEMENTS

First and foremost, I am extremely grateful to my two advisors, Dr. Ferath Kherif and Prof. Zoltán Kutalik. It has been a very fruitful and happy time for me to work with them. Both of them have motivated me with challenging and stimulating scientific questions, taught me how to think critically and showed me how to properly perform analysis and report my results. I appreciate all their technical and mental support in helping me to overcome all difficulties during my PhD period. I want to thank Ferath, who is also the subproject leader of the Human Brain Project, for the funding he provided to me.

I am thankful for the time and valuable comments from my thesis examiners, Prof. Giovanni Frisoni (UNIGE) and Mr. Jérôme Dauvillier (SIB). I enjoyed my defense and our interesting conversations. Many thanks also to the president of my thesis committee, Prof. Jean-Pierre Hornung (UNIL).

I am especially grateful to Prof. Richard Frackowiak. I have received a great help from him for my research projects. His creative thinking and enthusiastic attitude for all scientific areas has inspired me and showed me an excellent example of a successful professor. His efforts and encouragement have made my projects possible. As a charming gentleman, his sense of humor is very attractive and made all our communications a unique and pleasant experience.

The members of Laboratoire de Recherche En Neuroimagerie have been a source of good advice, friendship and fun. I want to also thank to Prof. Bogdan Draganski for his clinical expert suggestions in all my projects. I also thank Dr. Lester Melie-Garcia, for his great mathematical advice. Apart from that, I really enjoyed what he has shared with me his personal priceless stories. The time we have been working together is a good memory for my personal life and my career. I would also like to thank Antoine Lutti and Marzia De Lucia for sharing their suggestions.

I would like to thank my colleagues, Sandrine Muller, Valérie Zufferey, Juergen Dukart, Elisabeth Roggenhofer, Bart Van Damme, Gretel Sanabria-Diaz, Linda Dib and Anne Ruef and Aaron Mcdaid who have been contributed to my project, identification of preclinical cognitive normal people at risk of Alzheimer's disease. I would also like to thank my colleagues, Lester Melie-Garcia who has been contributed to my project, estimation of disease severity.

I thank my lab mates from LREN, Elham Barzegaran, Elsa Juan, Borja Rodriguez-Herreros, Lucien Gyger, Leyla Loued-Khenissi, Sara Lorio, Renaud Marquis, Sandra Martin, Claudia Modenato, David Slater, Maya Jastrzębowska, Estelle Dupuis and Remi Castella for beautiful memories of BBQ near Geneva lake, interesting games we played together, and for all the fun we have had in the last four years.

I also had a good time with fellows from SGG, Aurélien Macé, Sina Rüeger, Aaron McDaid, David Lamparter and Eleonora Porcu. They are curious, interesting and kind. I learned a lot from them on how to think critically and give constructive feedbacks. We also had a lot of fun from escape games and from exploring restaurants in Lausanne.

I had a great working experience with a special and brilliant team from HBP, including Mihaela Damian, Thanh Luu-Tho, Mirco Nasuti, Tea Danelutti, Arnaud Jutzeler and Ludovic Claude. We had a amazing collaboration and I enjoyed working with everyone from the team.

During my PhD study, I have received a great support from my family in China and Switzerland. A special thanks to my mother-in-law, Pauline, for her valuable effort in helping me to correct my papers and my whole thesis. I would also want to give a special thanks to my husband, Guillaume, who has given up many things for my research and career. Finally, I would like to thank to my parents for their support and encourage until now. I will always appreciate it.

ABSTRACT

Current Alzheimer's disease (AD) clinical diagnostic criteria rely on symptoms, which do not precisely reveal the underlying AD biological changes. Therefore, the criteria lack the ability to identify preclinical cases and quantify the disease severity objectively. I proposed new methodologies which use morphological markers to improve diagnostic accuracy. The rationale is that neuroimaging provides a gateway of integrating molecular and clinical markers.

My first study summarized the latest findings to outline AD biological mechanisms. We presented identified associations of neuroimaging features with clinical symptoms, and then molecular markers related anatomical and functional changes.

In my first experiment, I developed a classifier based on brain morphological markers enhanced with pathological information to identify individuals who are asymptomatic but at risk for clinical AD. This automated classifier was built using structural MRI of clinically and pathologically diagnosed subjects. This classifier was applied to predict pathology on cognitive normal (CN) subjects from two independent cohorts. Subjects predicted to have pathology showed significantly greater chance in clinical conversion to mild cognitive impairment or clinical AD, compared with those predicted negatively in both cohorts. Such difference was detected after taking into account of effects of other AD risk factors, like genetic variants. Among identified AD genetic variants, only APOE- ϵ 4 showed significant association with the clinical conversion. At the group level, predicted positive pathology group had significant atrophy in areas preferentially affected by AD except the hippocampus. Atrophy expanded to the hippocampus at a later stage.

My second project tested whether a latent trait estimated based on morphological markers can be used to quantify disease severity. MRI scans from two independent cohorts were used and their regional volumes were extracted. A latent variable model was implemented to estimate the latent trait based on the estimated regional volumes. The estimated latent trait had significantly different distributions between the clinical diagnosed groups and significantly associated with clinical conversion, from asymptomatic to symptomatic. The estimated latent trait significantly correlated with CSF proteins levels, hypometabolism and APOE- ϵ 4 genotype.

In conclusion, our results supported that morphological markers can identify cases at preclinical stages and quantify disease severity. Our methods might be useful to improve diagnostic accuracy in clinical decision-making.

RÉSUMÉ

Aujourd'hui, les critères du diagnostic clinique de la maladie d'Alzheimer (MA) sont basés sur des symptômes qui ne révèlent pas précisément de changements biologiques. Par conséquent, ces critères ne peuvent pas être utilisés pour identifier les patients précliniques et quantifier objectivement le stade de la maladie. Je propose les nouvelles méthodes qui utilisent des marqueurs morphologiques pour améliorer la précision du diagnostic. La raison sous-jacente est que la neuro-imagerie fournit une porte pour l'intégration des marqueurs moléculaires et cliniques.

Mon premier projet a résumé les dernières découvertes soulignant les mécanismes de MA. Nous avons présenté des liens établis entre les caractéristiques neuro-imageries et des symptômes cliniques, mais aussi de ceux-ci sur des modifications anatomiques et fonctionnelles relations avec les marqueurs moléculaires.

Dans ma première expérience, j'ai développé une classification automatique basée sur les marqueurs morphologiques cérébraux avec l'information de cette pathologie pour identifier les individus asymptomatiques, mais à risque. Cette classification automatique a été construite en utilisant l'Imagerie par Résonance Magnétique (IRM) structurelle de sujets cliniquement sains ou de patients atteints de MA. Cette classification a été appliquée pour prédire cette pathologie sur des sujets sans symptômes cognitifs de deux cohortes différentes. Les sujets qui ont été prédits comme pathologiques ont une probabilité plus grande d'avoir la MA, en comparaison des autres sujets. De tels différences ont été détectés après avoir pris en considération les effets d'autres facteurs de risques de la MA, comme différents effets génétiques. Parmi les effets génétiques seulement APOE- ϵ 4 montra une association significative avec le développement clinique. Les sujets qui ont été prédits comme pathologiques ont une atrophie significative dans l'aire affectée préférentiellement par la MA mais pas dans l'hippocampe. Cette atrophie s'étend à l'hippocampe à un stade plus avancé.

Ma seconde expérience a montré une variable latente estimée extraite des marqueurs morphologiques qui peut être utilisée pour quantifier le stade de la MA. J'ai utilisé les IRMs des deux différentes cohortes pour extraire ces marqueurs morphologiques. Un modèle multivarié a été mis en œuvre pour estimer la variable latente basée sur les marqueurs morphologiques. La variable latente estimée a une distribution différente entre les différents groupes et est associée de façon significative au changement de diagnostic clinique, comme de passer du groupe asymptomatique au groupe symptomatique. Cette variable latente est en corrélation avec le niveau de protéines dans le liquide céphalo-rachidien, l'hypométabolisme neuronal et APOE- ϵ 4 génotype, de façon significative.

En conclusion, nos résultats appuient que les marqueurs morphologiques peuvent identifier les cas précliniques et quantifier le stade de la MA. Notre méthode peut être utile pour améliorer la précision du diagnostic dans la prise de décision clinique.

LIST OF ABBREVIATIONS

AD	Alzheimer's disease
LOAD	late onset AD
CN	cognitive normal
MCI	mild cognitive impairment
SMC	subjective memory complaints
A β 42	amyloid beta (42-residue-long amyloid- β)
pTau	phosphorylated Tau
tTau	total Tau
CSF	cerebrospinal fluid
SNP	single nucleotide polymorphism
APOE	apolipoprotein E
APP	amyloid precursor protein
NFT	neurofibrillary tangles
NINCDS-ADRDA	national institute of neurological and communicative disorders and stroke and the Alzheimer's association
MMSE	mini-mental status examination
CDR	clinical dementia rating
FDG	fludeoxyglucose
PIB	pittsburgh compound B
PET	positron emission tomography
NIA-AA	national institute on aging-Alzheimer's association
GWAS	genome-wide association study
GRS	genetic risk score
SPM	statistical parametric mapping
VBM	voxel-based morphometrics

FWER	family-wise error rate
SVM	support vector machine
KM	Kaplan-Meier
Cox PH	Cox proportional hazards
HR	hazard ratio
PCA	principal component analysis
EBM	event-based model
MRI	magnetic resonance imaging
ASR	asymptomatic but at risk
CN_P+	cognitive normal predicted to have pathology
CN_P-	cognitive normal predicted to have no pathology
AD_P+	clinical AD predicted to have pathology
AD_P-	clinical AD predicted to have no pathology
ADNI	Alzheimer's disease neuroimaging initiative (America)
3C	three-city (France)
TIV	total intracranial volume
IRT	item response theory
FA	factor analysis
EM	expectation maximization
ML	maximum likelihood

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	I
ABSTRACT	III
RÉSUMÉ	IV
LIST OF ABBREVIATIONS	V
List of Figures	X
List of Supplementary Figures	X
List of Tables	XI
List of Supplementary Tables	XI
1. Introduction	1
1.1 Motivation	1
1.2 Main goal, methodology and challenges	2
1.3 Contribution	5
1.4 Thesis outline	6
2. Current Understanding of AD and Analytical Methods	8
2.1 Pathological mechanisms of disease	8
2.2 Clinical diagnostic criteria	10
2.3 Research diagnostic criteria	12
2.4 Analytical methods	12
2.5 Hill's causality inference	22
3. In-vivo brain neuroimaging provides a gateway for integrating molecular and clinical markers of Alzheimer's disease	23
3.1 Bridging neuroimaging biomarkers and clinical diagnostic	24
3.2 Bridging protein changes and brain anatomy and function	25
3.3 Bridging genetic variants and brain anatomy and function	28

3.4	Aging related biomarker changes	30
3.5	Conclusion	31
4.	Automated detections of asymptomatic individuals at risk for dementia due to Alzheimer's disease	33
4.1	Motivation	33
4.2	Introduction	33
4.3	Method	35
4.4	Results	38
4.5	Discussion	48
5.	A data-driven model of estimation a trait to quantify sporadic Alzheimer's disease severity	52
5.1	Motivation	52
5.2	Introduction	53
5.3	Method	54
5.4	Results	58
5.5	Discussion	65
6.	General Discussion	68
6.1	Findings of the projects	68
6.2	Causality inference	69
6.3	Methodological discussion	71
6.4	The extension of the three principles	74
7.	Reference	78
8.	Appendix	89
8.1	MRI preprocessing	89
8.2	Genetic variant processing	89
8.3	Linear FA with expectation maximization (EM) algorithm	89
8.4	Item characteristic function	90
8.5	Supplementary Figure	91

List of Figures

Figure 1 Representation of pathology in normal and AD patients.....	2
Figure 2 Biomarkers temporal evolution	9
Figure 3 One example of a SNP from one subject	13
Figure 4 Pipeline of neuroimaging preprocessing.....	17
Figure 5 Illustration of support vector machine algorithm	19
Figure 6 Example of survival data	20
Figure 7 Multiple data modalities in Alzheimer’s disease.....	24
Figure 8 Cerebral blood volume reductions patterns	26
Figure 9 Regional patterns of TDP-associated grey matter volume loss across Braak stages IV to VI.....	28
Figure 10 Cortical atrophy over 2 years in non-demented elderly adults	31
Figure 11 KM survival curves of predicted CN groups to MCI or AD in ADNI and 3C cohorts	44
Figure 12 Atrophy pattern across groups from PPG and ADNI cohorts	45
Figure 13 Atrophy pattern across groups from ADNI and 3C cohorts	46
Figure 14 Distributions of biomarkers at baseline between ADNI predicted groups.	47
Figure 15 MMSE distributions in follow-up study between ADNI predicted group.....	48
Figure 16 Association between the latent trait and clinical measurements in ADNI cohort ...	60
Figure 17 Association between the latent trait and biomarkers in ADNI cohort.....	63
Figure 18 Box plot of the estimated latent trait for APOE-ε4 groups.....	64
Figure 19 Regions strongly associated with the estimated latent trait	65

List of Supplementary Figures

Supplementary Figure 1 Volumetric difference between subgroups in hippocampal regions and entorhinal gyri	91
Supplementary Figure 2 Clinical score distributions in follow-up study between ADNI predicted group.....	92
Supplementary Figure 3 Values of 114 regions in factor loading matrix	94
Supplementary Figure 4 Expected regional ratios as a function of the latent trait values	95
Supplementary Figure 5 Adjusted volume of one voxel of amygdale across the estimated latent trait.....	96
Supplementary Figure 6 Expected regional ratios as a function of the latent trait values	97
Supplementary Figure 7 The estimated severity from FA compared with SVM predictions for ADNI subjects	98

List of Tables

Table 1 Recent findings about AD pathology-related biomarkers.....	11
Table 2 Explanation of Hill's criteria with smoking study as an example.....	22
Table 3 PPG, ADNI and 3C cohorts baseline demographic information.....	35
Table 4 Predicted groups of ADNI and 3C cohorts baseline demographic information ^a	41
Table 5 Effect of risk factors on conversion from cognitive normal to MCI/AD in ADNI and 3C cohorts.....	42
Table 6 Effect of each SNP on conversion from cognitive normal to MCI/AD in ADNI cohort.....	43
Table 7 ADNI and 3C studies baseline demographic information.....	61
Table 8 Effect of risk factors on conversion from cognitive normal to MCI/AD and from MCI to AD in ADNI and 3C cohorts.....	62
Table 9 Summary of two experiments results for testing a causality relation between atrophy and clinical AD conversion.....	70

List of Supplementary Tables

Supplementary Table 1 Effect of risk factors including CSF proteins on conversion from cognitive normal to MCI/AD and from MCI to AD in ADNI and 3C cohort.....	100
---	-----

1. Introduction

1.1 Motivation

With an increasing worldwide aging population, it has been predicted that a growing number of elderly people are going to be affected by neurodegenerative disease. The World Health Organisation (WHO) has reported that the number of people having neurodegenerative disease will reach 75 million by 2030 and 131 million by 2050. This increase will effect particularly low to middle income countries (Prince M, Wimo A, Guerchet M, 2015). Amongst the neurodegenerative cases, 50%-70% will be Alzheimer's Disease (AD) patients (Winblad *et al.*, 2016). As a result, AD will become the most common neurodegenerative disorder. It is becoming a high socio-economic burden for global economies.

Since the understanding of AD mechanism is limited, there is currently no cure to stop the progression of the disease. Several medications such as cholinesterase and acetylcholinesterase inhibitors have been demonstrated to be useful in delaying AD symptoms for a limited period of time (Scheltens *et al.*, 2016).

The above drugs can delay the progression but they cannot stop it. New drugs are aiming to either increase the clearance rate of suspected neurotoxic markers or to reduce their neurotoxicity. The neurotoxic markers (such as amyloid- β , Tau and phosphorylated Tau (pTau)) are proteins that can form plaques and neurofibrillary tangles in the brain of patients (see Figure 1) and they have been long been the prime suspect for causing the disease (Jucker and Walker, 2013). Ongoing research and development of a cure is focusing on anti-amyloid drugs such as the antibody aducanumab (Sevigny *et al.*, 2016). Aducanumab has been recently shown to benefit patients by reducing amyloid levels in the brain in an ongoing clinical trial. Such anti-amyloid drugs have been considered to be most effective in patients with early Alzheimer's disease (Scheltens *et al.*, 2016). The detection of Alzheimer's disease at early or even preclinical stages becomes key to provide the patient with early treatments to sustain their quality of life.

Current clinical AD diagnostic criteria mainly rely on symptoms and lack the power to reveal pathological brain changes. It is necessary to identify people at risk of developing clinical AD, which is when the symptoms (memory loss, decline in cognitive performance) manifest themselves and dementia is present.

Several markers have been tested for their diagnostic usage. Such markers include proteins levels from cerebrospinal fluid (CSF) / cortical regions (extracted from molecular neuroimaging). The drawback of the above methods is that they are expensive (molecular neuroimaging) and invasive (CSF). The

main difficulty faced by researchers is to integrate low-cost and non-invasive methods in a convenient way to enable clinicians to diagnose AD, especially at the preclinical stages.

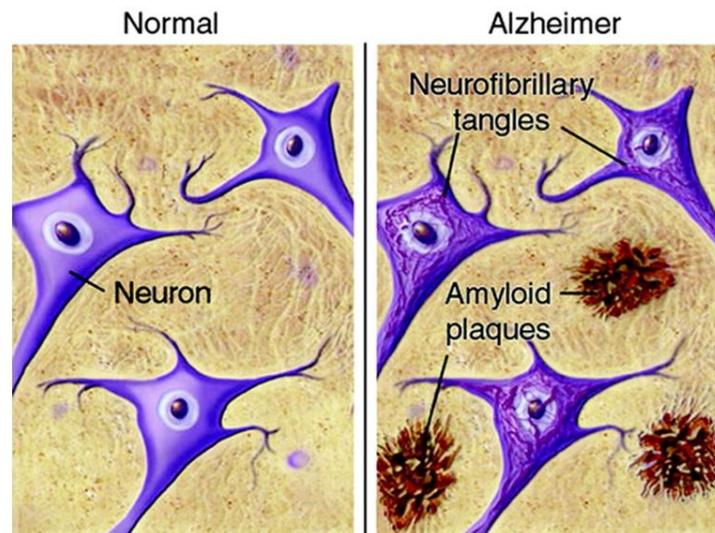


Figure 1 Representation of pathology in normal and AD patients.

This illustration shows tissues from normal and AD patient brains. Compare to normal brains, The patients have propagations of proteins (tangles inside of brain cells and plaques in between the brain cells). This figure was extracted from (Silbert, 2007).

Identifying preclinical cases is important for early prevention but a continuous disease severity trait is more useful to quantify disease stages. A continuous estimation of disease severity is more useful since Alzheimer's disease, like other types of neurodegenerative diseases, is a progressive brain disorder. An estimator of disease severity should have the ability to quantify different stages and monitor dementia onset. The definite staging of the disease can only be confirmed in an autopsy examination. The difficulty is to be able to estimate AD severity *in vivo* that best reveal the pathological changes in the brain using low-cost, convenient and non-invasive measurements.

1.2 Main goal, methodology and challenges

The main goal of this thesis is to build disease models based on the features derived from MRI and interacting information from other biomarkers and clinical phenotypes. The objectives are to improve to accuracy of AD, to predict cognitive normal subjects who are at risk to develop clinical AD and quantify AD severity *in vivo*.

To reach this goal, in this thesis, we develop a methodological framework based on 3 principles: computational anatomy, disease modeling and biomarkers interaction used for construction and validation models. The 3 principles and related challenges are explained in the paragraphs below.

Principle 1: Computational anatomy

In my experiments, I relied on computational anatomy methods developed for neuroimaging research. Computational anatomy extracts morphological features from MRI scans, by estimating volumes of brain tissues in an automated way (Ashburner, 2009). Computational anatomy has been widely used to identify atrophy patterns of mild cognitive impairment (MCI) and AD (Chételat *et al.*, 2005; Karas *et al.*, 2004; Krasuski *et al.*, 1998) to identify diseases linked anatomical structures differences. However, there are still several challenges associated with this method for building a reliable individual-based disease model.

Challenges:

- Limitations of group comparison and unreliable diagnosis: The applications of the previous methods for identifying AD atrophy often rely on labels from clinical diagnosis. However, clinical diagnoses are known to be unreliable in neurodegenerative diseases (Sperling *et al.*, 2011). Therefore, group comparison between clinical diagnosed cases and controls does not precisely reveal disease atrophy, especially at early stages. Moreover, the results are informative at group level but not at individual levels. The reason is not only due to the method itself but also because there is remarked inter-individual variability in gyral volume (Alexander-Bloch *et al.*, 2013).
- Univariate: Univariate methods analyze one variable at a time, such as one voxel or region-of-interest. For example, several studies used regional volumetric features, such as hippocampal volumes (Coupé *et al.*, 2015; Jack *et al.*, 2014; Stephan *et al.*, 2015). The drawback of only using regional volumetric features is that the spatial interaction of different regions is ignored and therefore the information is insufficient (Frisoni *et al.*, 2010).
- Confounder: There are several factors that can affect the estimation of volumetric features, including intrinsic factors like age and brain size and technical factors like recruiting procedure. Their effects can be reduced by integrating those confounders in a statistical model. But it is not possible to control their effects in a multivariable model, for instance in differentiating groups (Klöppel *et al.*, 2008).

Principle 2: Disease modeling

I have constructed disease models using machine learning algorithms. In contrast to univariate method which focuses on only one variable, multivariable model can integrate all volumetric features or multiple variables of interest. Such a model is useful to predict outcomes or detect disease-related or outcome-related variables at the individual level. Through a multivariable model, effect of one variable of interest can be determined after taking into account other confounding variables. Thus,

confounder effects can be modeled and adjusted. Disease models built to predict outcomes have been shown in several studies of neurodegenerative diseases (Bron *et al.*, 2014; Klöppel *et al.*, 2008; Lillemark *et al.*, 2014; Peters *et al.*, 2014).

Challenges:

- High dimensionality: Neuroimaging data has high dimensionality as the number of volumetric features is numerous. The number of subjects is much smaller than the number of dimension in most of the studies. A strategy that avoid over-fitting can be helpful.

Principle 3: Biomarkers interaction used for construction and validation models

As mentioned above clinical diagnoses are unreliable and univariate methods are insufficient for AD. Integrating information from other modalities can increase the reliability of a disease model. In my experiments, I integrated information from other modalities for constructing and validating my disease models. Using them as inputs of my model, I have integrated the “gold standards” diagnosis with clinical diagnosed information to overcome the limitation of clinical diagnosis. In validation, a good disease model is suspected to represent the biological process of the disease. Since AD pathology is inaccessible *in vivo* and the mechanism is unknown, the validation of my models can be based on discovered interactions identified.

For the validation of given volumetric features, its interactions with AD pathology are based on identified evidence. Volumetric features correlate strongly with the distribution of neurofibrillary tangles (Frisoni *et al.*, 2010; Braak and Braak, 1995; Giannakopoulos *et al.*, 2003; Whitwell *et al.*, 2008, 2012).

In vivo, accessible proteins levels (CSF protein levels, molecular imaging detected cortical protein burden) may be useful, as several studies have demonstrated their correlations AD pathology (Buerger *et al.*, 2006; Strozyk *et al.*, 2003; Tapiola *et al.*, 2009). Validating a model can be done by testing its relation with these accessible markers to examine whether such a model mirrors biological changes. In addition, the model can be tested for its predictive power of clinical AD beyond these accessible markers. Such markers include proteins mentioned above and genetic variants associated with clinical AD, which termed AD genetic risk factors.

Challenges:

- Small effect variables: For some variables, their impacts on disease are difficult to be proven. Such as single nucleotide polymorphism (SNP), the effect size of each variable is mildest. There is a need to increase their ability for disease prediction.

1.3 Contribution

In this thesis, we bring to light three contributions:

1.3.1 Brain neuroimaging provides a gateway for linking molecular and clinical markers

My first work reviewed identified associations between different modalities to understand AD disease mechanisms from currently published studies (published on *Current Opinion in Neurology*, 2015)(Cui *et al.*, 2015). This work provides evidence for the 1st and 3rd principles, “computational anatomy” and “biomarkers interaction used for construction and validation models”. We demonstrated that neuroimaging can be used as a gateway of linking molecular markers and cognitive symptoms.

1.3.2 Identification of asymptomatic individuals at risk of AD conversion

My 1st experimental project was based on the 3 principles from the methodological framework described in section 1.2. I applied a classifier (2nd principle), which is capable of predicting AD pathology, on cognitive normal subjects and identify asymptomatic individuals at risk of developing AD (3rd principle). Pathology-related volumetric features were extracted from MRI scans as the input of a supervised learning algorithm to construct this classifier to predict pathology (1st principle). Before constructing this model, the volumetric features were normalized and then the effects of a set of confounders have been modeled and removed. These steps were done to overcome challenge “confounder”. The integration of pathology information and volumetric features provides a solution to two challenges described above, “limitations of group comparison and unreliable diagnosis” and “univariate”. In the process of constructing the model, I have used a linear kernel to overcome the “high-dimension” challenge. The classifier was used to predict pathology among cognitive normal subjects. The predicted pathological states showed a significantly association with clinical conversion in the follow-up study. For variables with small effect size (“small effect variables” challenge), like SNP, a genetic score has been calculated based on SNPs of interest to test an overall effect.

This is the first study to apply a classifier to predict pathology on cognitive normal participants to identify a group of healthy people who are at risk to develop clinical AD. The results provided further evidence that pathology-related volumetric change occurs before the symptoms, even among cognitive normal people. Such changes significantly associated with clinical AD conversion. The application of this experiment may improve clinical decision-making to detect the group might be benefited most from early treatment.

1.3.3 Estimation of disease severity based on biological data

The 2nd experiment is an extension of the 1st one. Learning that Alzheimer's disease is a progressive disorder and pathology occurs before symptoms onset, I extracted a continuous index from brain anatomical features (1st principle) by applying a data-driven method (2nd principle). Before using the data-driven method, I applied a scaling method and then the effects of a set of confounders have been modeled and removed. These methods can be used as an alternative solution to challenge "confounder". Because the data-driven method used in this experiment does not require the group labels of each subject, therefore such method provides an alternative solution to the same challenges mention in section 1.3.2, "limitations of group comparison and unreliable diagnosis" and "univariate". The extracted continuous index demonstrated that it can be used to quantify disease severity and monitor symptoms onset, after comparing with pathology-related markers level and clinical measurements at baseline and follow-up study (3rd principle).

This experiment is the first one applied to a data-driven method using continuous measurements to detect disease-related traits. These disease-related traits have been shown to be significantly correlate with pathology-related proteins, extracted from CSF and cortical regions. This method can be used in clinical practice to assist the quantification of disease severity and monitor the risk of symptoms onsite for cognitive normal and MCI people.

1.4 Thesis outline

The thesis is structured according to the three main contributions outlined in section 1.3. Each comprising a chapter and the following sections give an outline of each chapter.

Chapter 2 Current Understanding of AD and Analytical Methods

In this chapter, we first introduce the proposed disease mechanisms, including amyloid- β and Tau hypotheses, from previous studies. Then, AD clinical and research criteria are compared and the main difficulties we are facing in the experiments are listed. Machine learning algorithms and statistical methods used in my experiments as well as methods relevant are all explained. The last section presents Hill's causality criteria, which will be used in Chapter 6.

Chapter 3 In-vivo brain neuroimaging provides a gateway for integrating molecular and clinical markers of Alzheimer's disease

In this chapter, recently identified interactions between multiple modalities are summarized. Modalities include genetic variants, proteins, brain anatomy / function and clinical measurements. This review covers interactions between modalities.

Chapter 4 Automated detections of asymptomatic individuals at risk for dementia due to Alzheimer's disease

In this chapter, we present the steps we took to build an automated classifier using volumetric features extracted from MRI of clinically and pathologically diagnosed subjects. We summarize the results of applying this classifier which predicts pathology on cognitive normal subjects from two independent cohorts. We list results from testing the relative risk of developing clinical AD between predicted positive and negative pathology groups, compared with other AD risk factors, such as age and genetic variants. Anatomical differences between the two predicted groups compared at baseline and in the follow-up study are illustrated.

Chapter 5 A data-driven model of estimation a trait to quantify sporadic Alzheimer's disease severity

In this chapter, we extend the idea of chapter 4 to estimate a continuous trait rather than a binary classification for each subject. We describe a data-driven method using volumetric features to estimate a continuous trait for two independent cohorts. By utilizing markers, including clinical measurements (at baseline and in follow-up) and molecular markers, we demonstrate this trait can be used to quantify disease severity. Meanwhile, we show relations between the estimated trait and grey matter regions.

Chapter 6 General Discussion

We conclude by providing a holistic summary of our work. The causal inference of our models and symptoms onset is explained. For the challenges mentioned in section 1.2, the solutions and the interpretations are provided. We draw attention to the current open problems and directions for future work in the field of understanding the neurodegenerative diseases.

2. Current Understanding of AD and Analytical Methods

2.1 Pathological mechanisms of disease

Alzheimer's disease, named after the German doctor Alois Alzheimer, is defined as a disease with memory impairment prior to death and a dramatic shrinkage in the brain with abnormal deposits, revealed at autopsy. These abnormal deposits are known to be amyloid plaque and neurofibrillary tangles.

Two hypotheses, amyloid and Tau hypothesis have been proposed as the mechanism of Alzheimer's disease.

The amyloid hypothesis, proposed for more than ten years, considers the deposition of amyloid- β in amyloid plaques in brain tissue as the cause of the neurodegeneration in Alzheimer's disease (Hardy and Selkoe, 2002). The 42-residue-long amyloid- β (A β 42) peptide rapidly aggregates to form oligomers, which are the main components of plaques (Ahmed *et al.*, 2010). Aggregated A β 42 affects neurons and induces neurofibrillary tangle formation that leads to neuronal death and impairment of synaptic plasticity (Querfurth and LaFerla, 2010; Takashima, 2009). A β 42 peptides are derived from amyloid precursor protein (APP) by beta-site amyloid precursor protein-cleaving enzyme 1 and γ -secretase (Querfurth and LaFerla, 2010). Experimental application shows that high concentrations of A β 42 oligomers suppress basal synaptic transmission by facilitating endocytosis of receptors of neurotransmitters. In severe late stage AD, patients have a reduced level of neurotrophin receptors due to the binding of A β 42.

In recent research Tau hypothesis shows a closer correlation between the loss of neuronal function and the degree of neurofibrillary tangle (NFT) rather than the degree of senile plaque accumulation (Cripps *et al.*, 2006). It has been recently shown that hyperphosphorylated Tau is the major protein component of NFTs in AD (Cripps *et al.*, 2006). The abnormal phosphorylation of Tau leads to its dissociation from microtubules and the formation of NFTs (Giannakopoulos *et al.*, 2003). These NFTs are highly accumulated in the entorhinal and hippocampal neurons and correlate with the memory impairment observed in AD (Giannakopoulos *et al.*, 2003). The onset of Tau modification leads to sequential neuronal dysfunction, starting with synaptic loss and impairment in axonal transport which leads to NFTs formation and neuronal loss, see Figure 2 (Jack and Holtzman, 2013; Takashima, 2009).

AD can be either in dominantly inherited or sporadic form with different age of onset. Less than 1% of AD cases are autosomal dominant AD, referred to as familial AD, due to the mutation of 3 genes, APP, presenilin-1 and presenilin-2. Most of familial AD cases occur before 65 years old. Mutated Tau evolves into multiple neurodegenerative diseases, such as frontotemporal dementia. In Alzheimer's disease, Tau mutations do not occur (Goedert and Jakes, 2005). The life-time risk for sporadic AD is

appearing from the age of 65 on average and leads to death after about 10-15 years (Dubois *et al.*, 2016).

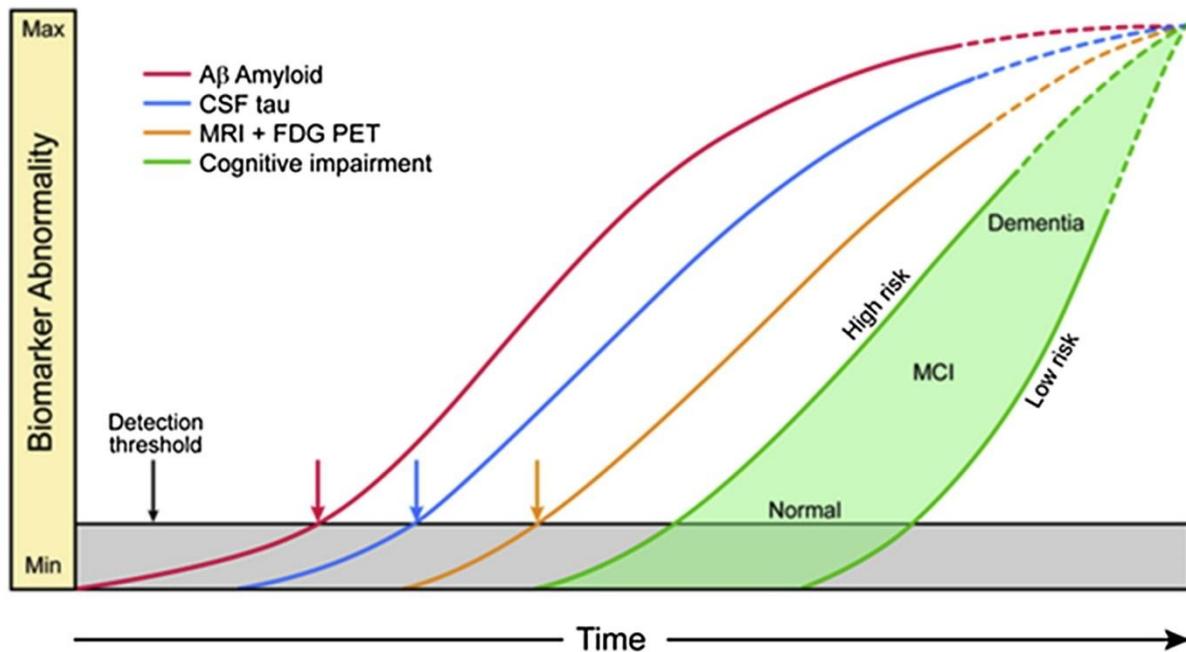


Figure 2 Biomarkers temporal evolution

This figure shows hypothetical typical AD biomarkers change sequence. The x axis represents time and y axis represents a change from normal to abnormal levels. The grey line indicates the detection threshold level for biomarkers. In typical AD, A β is supposed to change first and then followed by CSF Tau protein. MRI features and hypometabolism change afterward. Cognitive symptoms appear at the end. This figure is adapted from (Jack and Holtzman, 2013).

AD progression has been divided into six stages by Braak and Braak. Stages I-II show alterations virtually confined to the transentorhinal region (temporal lobe) and named as “transentorhinal stages”. At these stages, NFTs slowly build up at their predilection sites, but is asymptomatic. Stages III-IV are characterised by an important involvement of both the entorhinal and transentorhinal regions and the hippocampus accompanied by large numbers of Pre α projection cells carrying NFTs. These two stages are also called “limbic stages” and clinical protocols can be used. At the neocortical stages, stages V-VI, there is destruction of the inferior temporal and lateral temporal cortex and other neocortical association areas. This explains the impairment of other cognitive domains apart from episodic memory (personally experienced events in a particular temporal and spatial context) such as language problems, disorientation in time and place as well as motor dysfunction in the end stage of the disease (Braak and Braak, 1995).

2.2 Clinical diagnostic criteria

The current clinical Alzheimer's disease diagnostic criteria mainly rely mainly on symptoms (Scheltens *et al.*, 2016). The mostly applied criteria are NINCDS-ADRDA Alzheimer's Criteria (National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Association) (McKhann *et al.*, 1984). They specify eight cognitive domains that may be impaired in AD, including memory, language, perceptual skills, attention, constructive abilities, orientation, problem solving and functional abilities. In addition, two other measurements of cognitive performance are the mini-mental status examination (MMSE) and clinical dementia rating (CDR). Where the diagnosis is based on clinical measurements the patients are probable AD or possible AD. The definite diagnosis requires a further histopathological confirmation of AD by biopsy or autopsy.

Biomarkers	Measurements	Summary (References)
Aβ42	CSF	<ul style="list-style-type: none"> • Low Aβ42 was the strongest predictor of clinical progression in patients with subjective complaints (Van Harten <i>et al.</i>, 2013) • Aβ42 oligomers levels in CSF is clearly associated with AD, the overlap is too large (Hölttä <i>et al.</i>, 2013) • The prevalence of amyloid pathology increase with age among normal cognitive participants, subjective cognitive impairment and mild cognitive impairment patients. (Jansen <i>et al.</i>, 2015)
	Cortical	<ul style="list-style-type: none"> • The likelihood of amyloid positivity decrease with age among AD patients (Ossenkoppele, Jansen, <i>et al.</i>, 2015) but increase with age among non-AD dementias, like frontotemporal dementia, vascular dementia and Lewy bodies patients (Ossenkoppele, Jansen, <i>et al.</i>, 2015)
Total Tau	CSF	<ul style="list-style-type: none"> • CSF total tau and pTau correlate with AD. Total tau increases in traumatic brain injury, stroke and in Creutzfeldt-Jakob disease, but not elevated CSF pTau levels (Jack <i>et al.</i>, 2016).
pTau	Cortical	<ul style="list-style-type: none"> • ¹⁸F-AV-1451 positron emission tomography can be used to accurately quantify in vivo the regional distribution of hyperphosphorylated tau protein among patients with MAPT R406W mutation (Smith <i>et al.</i>, 2016) • the pattern of increased [¹⁸F]AV-1451 highly overlapped with regions that showed decreased [¹⁸F] Fludeoxyglucose (FDG) uptake (hypometabolism), [¹¹C] Pittsburgh compound B (PIB) binding (cortical Aβ42) showed no correlation with [¹⁸F]FDG uptake (Ossenkoppele, Schonhaut, <i>et al.</i>, 2015)
CSF Aβ42 and tau concentrations and a memory composite score		<ul style="list-style-type: none"> • Preclinical Alzheimer's disease exists in 30% cognitively normal elderly people and it associated with future cognitive decline and mortality (Vos <i>et al.</i>, 2013)

Table 1 Recent findings about AD pathology-related biomarkers

2.3 Research diagnostic criteria

Pathology related biomarkers are the measurements of protein A β 42 and Tau levels in CSF and in the brain. The cortical amyloid burden can be directly detected by positron emission tomography (PET) imaging. Currently the most widely used tracer to image cortical A β 42 is 11C-labeled Pittsburgh compound B (PIB) PET and florbetapir F 18 (also known as [18F] AV-45) PET. Apart from measuring cortical A β 42 burden using PET imaging directly, cerebrospinal fluid (CSF) A β 42 levels (Strozyk *et al.*, 2003) and phosphorylated Tau (pTau) protein at threonine 231 (pTau231) and threonine 181 (pTau181) have been demonstrated to reflect pathologic changes in the brain (Tapiola *et al.*, 2009),(Buerger *et al.*, 2006) and are associated with longitudinal clinical conversion (Dubois *et al.*, 2016). Although pTau231 protein shows a better correlation with neurofibrillary tangle burden at autopsy(Buerger *et al.*, 2007), pTau181 is widely studied so far. The ratio between CSF phosphorylated Tau threonine 181 and A β 42 (pTau181/A β 42) has been proposed as one of the best biomarkers to predict cognitive decline in nondemented older adults (Fagan *et al.*, 2007) and AD conversion risk in MCI group (Buchhave *et al.*, 2012). Latest findings of these biomarkers are summarized in Table 1.

As clinical diagnostic criteria of AD are unable to reveal the pathology, especially at an early stage, in 2007, Dubois et al proposed new criteria to better define Alzheimer's disease (Dubois *et al.*, 2007). Similarly, new criteria have also been proposed by National Institute on Aging–Alzheimer's Association (NIA–AA) in 2011 (Sperling *et al.*, 2011). Both research diagnostic criteria integrate biomarkers into the diagnostic process and therefore covered the full span of the disease, from preclinical to severe AD. The two diagnostic criteria differ in which are the selected biomarkers for different stages. At preclinical stage, only the NIA–AA criteria include MRI markers, IWG do not. In 2016, both criteria were merged for the preclinical stage and in the new version, only the pathology-related biomarkers are considered. However, this version has some issues when applied to identify from the general population those who are at preclinical stage of AD.

2.4 Analytical methods

2.4.1 Genome-wide association study

Three billion base pairs of genetic information in human are 99.9% identical between individuals. Among the 3 billion base pairs, there are around 10 million single nucleotide polymorphisms (SNP), which is one type of genetic variants. A SNP is a single DNA base-pair change at a specific base position in the human genome, with the possibility of presenting one of the 4 possible nucleotides (A, T, C and G). A nucleotide that presents at a SNP is called an allele (i.e G allele). Each person has two copies of the same gene inherited from each parent, one SNP has two alleles. A genotype for this SNP is defined by the alleles which one person has at this position.

When parents pass on their genetic information to their children there are blocks of nucleotides that are always inherited with little recombination forming linkage disequilibrium (LD) blocks, which contains dependent allelic variants, and give rise to haplotypes. As a result, from the original 10 million SNP variants there are in fact only 300,000-1,000,000 distinctive clusters (Hafler and Jager, 2005). Genome-wide association studies (GWAS) aim at identifying SNPs whose alleles correlate with traits or diseases in various groups of human subjects.

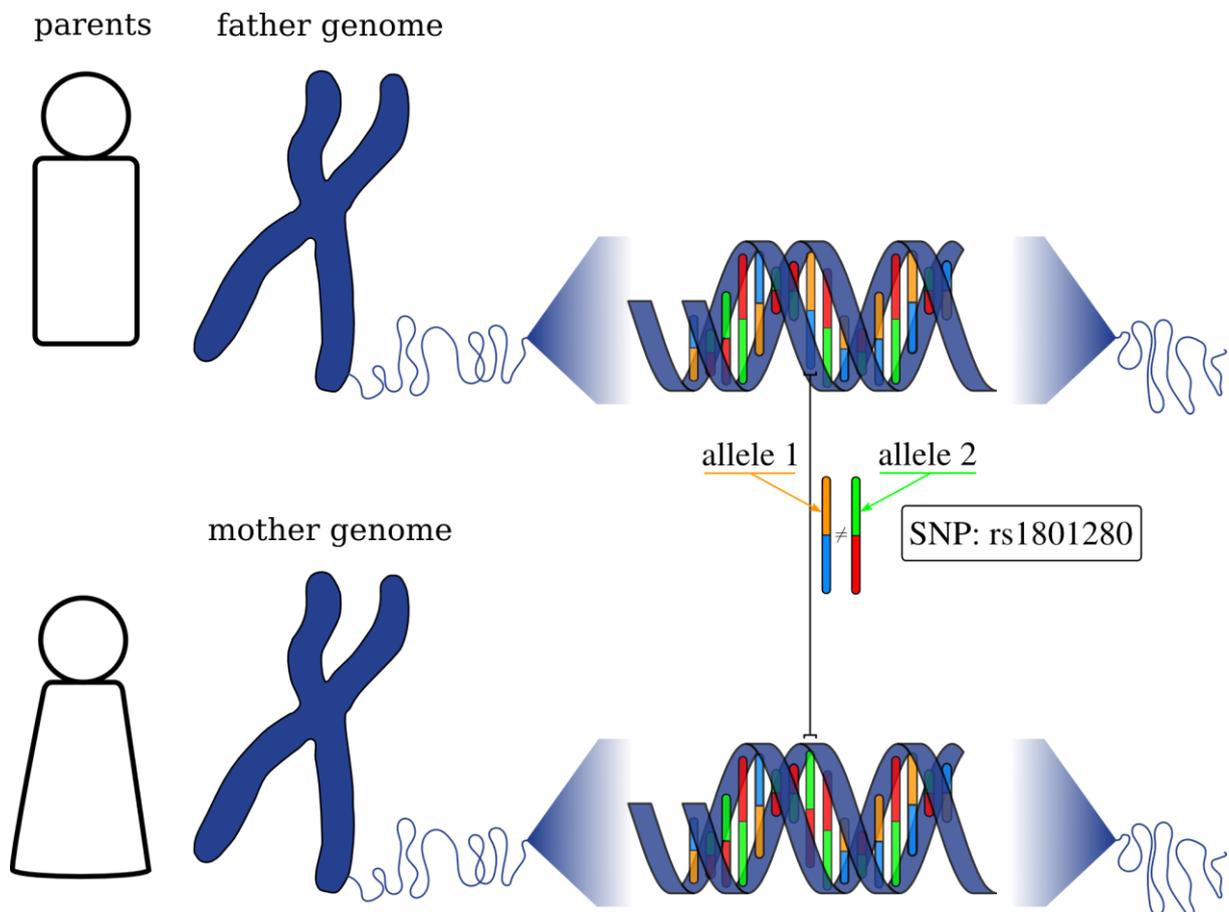


Figure 3 One example of a SNP from one subject

This figure shows an example of one SNP (i.e rs1801280) from one subject. SNP is a single base change in a DNA sequence. The two sequences represent the two transmitted copies inherited from the subject's parents. The genotype of the subject at this SNP is G/A. The subject has a minor allele (i.e G in green) inherited from the mother (for rs1801280), and a major allele A (same as the reference) from the father (in orange).

In GWAS, subjects and SNPs which do not fulfill specific quality criteria are removed in the different preprocessing steps (Anderson *et al.*, 2010). Subjects passing the following criteria are retained for analysis: no discordance between genetic and self-reported sex, low missing genotype rate (<5%), reasonable heterozygous rate (~25-30%) and unrelated to other samples. All subjects which do not

fulfill these conditions are not considered during the GWAS. Sexual discordant individuals have a mismatched sex according to their genotype data. Subjects with a high missing genotype or heterozygosity rate may arise be due to fabrication error of the chip or sample mixing, poor DNA quality, etc.

QC criteria for SNPs include: high call rate (>98%) and no differential missingness, >1% minor allele frequency and Hardy-Weinberg equilibrium (HWE) test ($p < 0.0001$).

SNPs with very low frequency of the minor allele should be removed as genotype clustering algorithms tend to be less reliable when some clusters have too little individuals. HWE is based on Hardy-Weinberg law that in a large random mating population, for a given SNP and its minor allele (a) frequency (p), the expected frequencies of the three unordered genotypes frequencies (aa, aA and AA) can be calculated ($p^2, 2p(1 - p), (1 - p)^2$). Deviation from HWE can be caused by several reasons. But most studies suggest it is an indication of population stratification rather than genotyping error (Cox and Kraft, 2006). SNPs show significantly deviations from the expected genotype frequencies are removed. To avoid disease bias, HWE test is normally done using a control group.

Once the two preprocessing steps are done (removing subjects and SNPs which do not meet the conditions outlined above), we proceed to the GWAS.

To identify disease-related genetic risks such as the risk of Alzheimer's the most widely used model is logistic regression which is a probabilistic binary linear classifier (Bush and Moore, 2012; Hastie *et al.*, 2009). Using Alzheimer's disease as an example, the probability of a subject having the disease $\Pr(G = disease|X = x; \beta)$, where $G \in \{disease, control\}$ are the phenotypes, X indicates the predictors (age, gender, SNP genotype and the first few principal components which corrects for the population structure) and β are the parameters of the model to be estimated. Logistic regression assumes that the logit of the disease probability, $\Pr(G = disease|X = x; \beta)$, is a linear function of the predictors, see equation (2.1).

$$\begin{aligned}
 &= \text{logit}(\Pr(G = disease|X = x)) \\
 &= \log\left(\frac{\Pr(G = disease|X = x)}{1 - \Pr(G = disease|X = x)}\right) \\
 &= \beta_0 + \beta_{age}X_{age} + \beta_{sex}X_{sex} + \beta_{SNP}X_{SNP} + \beta_{pc1}X_{pc1} + \beta_{pc2}X_{pc2} + \beta_{pc3}X_{pc3} \quad (2.1)
 \end{aligned}$$

The genetic variable $X_{SNP} \in \{bb = 0, bB = 1, BB = 2\} \in \mathbb{R}$ is numeric. The small b represents the reference allele and capital B represents the coded allele. The first letter in bb (as an example) represents the genotype one inherited from his or her mother and the second letter is from his or her father. The parameters of the logistic regression are learned via maximizing the log-likelihood.

There are three main options for encoding the genetic information X_{SNP} : additively, dominantly and recessively, depending on the different assumptions of the risk allele affecting the phenotype. The additive method assumes that the more risk allele a subject has, the greater the effect will be observed. For the additive method, the three types of allele (bb, bB, BB) are coded as 0, 1 and 2, as the number of risk allele increase, $X_{SNP} \in \{bb = 0, bB = 1, BB = 2\}$. The dominant method assumes that the effect of the phenotype can be seen already when one risk allele appears, $X_{SNP} \in \{bb = 0, bB = 1, BB = 1\}$. In contrast, the recessive model requires the two copies of the risk allele to be present to take an effect, $X_{SNP} \in \{bb = 0, bB = 0, BB = 1\}$.

The last three predictors in the logistic model ($X_{pc1}, X_{pc2}, X_{pc3}$) represent divergent ancestry, where $X_{pc1} \in \mathbb{R}$ is the first principal component score whilst X_{pc2} and X_{pc3} are the scores of the second and third principal component of the genetic kinship matrix. The idea is to use the genotypic features to estimate genetic kinship and through principal component analysis (PCA) identify clusters of subjects coming from the same ancestry. Individuals that are outliers ($>3SD$ away from the mean) in terms of principal component loadings are removed from further analysis. The first few (typically 2-10) principal components are included in the logistic model as covariates to reduce confounding due to population structure.

The GWAS method described above identifies novel genetic risk factors (SNPs) that are related to a given disease phenotype. This is achieved by learning a logistic regression model over the entire genome to detect the most discriminative SNPs.

2.4.2 Genetic risk score

Given a list of SNPs associated to a disease (identified in GWAS) one can calculate their weighted average, termed Genetic Risk Score (GRS), where the weights are essentially effect sizes obtained from the study.

For a complex disease several factors (predictors) may be involved such as genetic and environmental factors. For neurodegenerative disease sporadic cases, the observed phenotypes are likely to be the results of a mixture of different factors. Since the effect sizes of individual genetic variants are rather small it is reasonable to combine a set of identified disease related SNPs into one score.

There are two solutions to combine a set of SNPs. The first is to calculate the total number of risk allele each subject has. Equation (2.2) shows the calculation of GRS for subject i with J risk SNPs where n_{ij} is the number of risk allele of subject i at SNP j . The second solution is to weight each SNP effect and then sum them up. The weights are the log (odd ratios) of the SNPs identified from previous

studies. As shown in equation (2.3), OR_j is the odds ratio of the j th SNP. This method has been tested among diseases, such as psoriasis (Chen *et al.*, 2011).

$$GRS_i = \sum_{j=1}^J n_{ij} \quad (2.2)$$

$$wGRS_i = \sum_{j=1}^J (n_{ij} * \log(OR_j)) \quad (2.3)$$

2.4.3 Neuroimaging preprocessing

Morphometrics is a study of the variability of the form of organisms or objects (Ashburner, 2009). To be able to compare two brain MRI images, it is necessary to map them into one standard space. Statistical Parametric Mapping (SPM) is a software which is widely used in computational anatomy to perform voxel-based morphometrics (VBM) (Ashburner, 2009). This software extracts volumetric features from the shapes of MRI images. VBM requires several preprocessing steps. The preprocessing starts by assigning each voxel from T1-weighted MR scans into a particular tissue type: grey matter, white matter or cerebrospinal fluid, based on its intensity. This step is done via a rigid-body algorithm to roughly align the subject to a common anatomical space known as the tissue probability map. Then, each of the voxel is assigned into a tissue type with the highest probability estimated through a mixture of Gaussians model, which represents the intensity distributions of different tissue classes. Using grey matter is an example, the grey matter probability maps of all subjects will be used in an inter-subject registration algorithm, named as Dartel. Dartel starts by averaging all subjects' grey matter maps to generate a template and it is followed by an iterative procedure of minimizing the mean squared difference between the individual's image and the template. The outputs of Dartel are the warped images that align with the average-shaped template and flow field maps that encode how each grey matter probability map deformed to best match the template. However, the coordinates of the outputs of this step cannot be compared with other studies. One more spatial transformation step is needed so that the grey matter tissue class image of each individual can be aligned with the Talairach and Tournoux, or MNI space. To preserve the volume of each voxel, the deformed images are multiplied with the Jacobian determinants of the deformations from flow fields. It works in a way that if one region is expanded or shrunk compared to its original volume during deformation, and then the intensity would be reduced or enlarged to keep the signal of that region conserved. The Jacobian-corrected, warped tissue class images would then be smoothed to increase the signal to noise ratio. The smoothing step is done by convolving with an isotropic Gaussian kernel. The results of all the steps represent the regional volume of tissue and can be used for further statistical analysis in VBM.

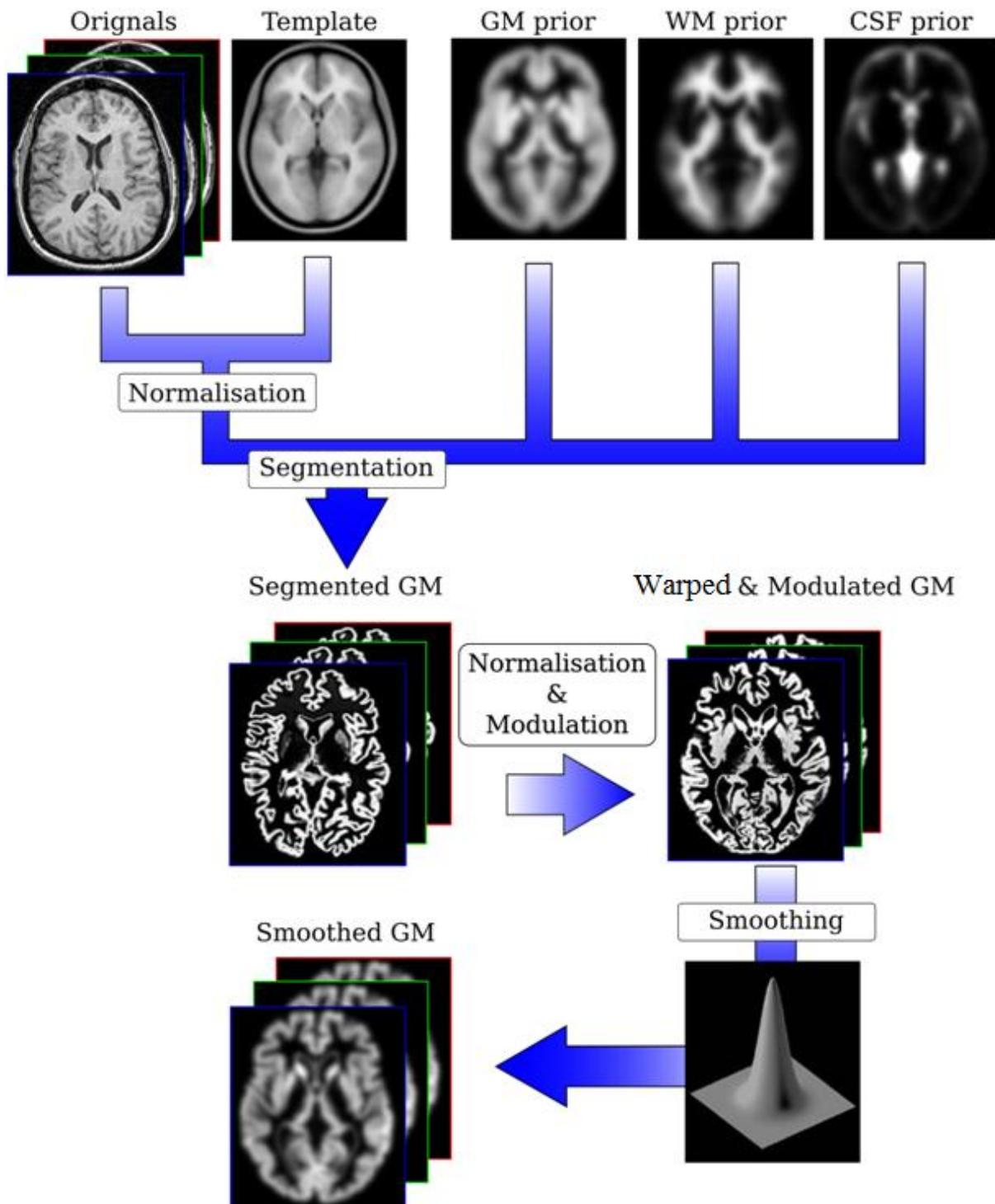


Figure 4 Pipeline of neuroimaging preprocessing

This figure shows all steps used in preprocessing structural neuroimaging data. It begins with segmentation, assigning each voxel of the scan to a tissue type, i.e grey matter. The segmented grey matter probability maps are then warped to a common space. The volume of each voxel is preserved at the mean time through modulation. The outputs are then spatially smoothed with a Gaussian kernel to increase the signal to noise ratio. The final results can be used in voxel-to-voxel statistical analysis.

VBM is a mass-univariate voxel-based morphometry method to investigate topographical, grey matter volume, difference at voxel level using the framework of the general linear model. The estimated volumes of all grey matter voxels are compared independently between the case and the control groups. In a general linear model, disease related factors are normally used as cofounders to be adjusted. For one voxel, the model is fitted through finding the coefficients of the covariates that minimize the residuals. To answer a question, such as what regions are significantly different between patients and controls, statistical tests (t test and F test) can be used to test whether the disease effect is significant from zero and proportion of total variance explained by the model is significantly different from zero.

2.4.4 Family-wise error rate

In both GWAS and VBM studies, genome-wide and voxel-wise models are built to identify phenotype-related markers. In a simple case of testing the association between a trait and one SNP or one voxel, the null hypothesis (H_0) is that there is no association between this SNP or this voxel and the trait. The alternative hypothesis is the complement of the null hypothesis. A statistical test can be performed and the test statistic with a corresponding p-value can be calculated. The p-value represents the probability of obtaining a more extreme test statistic than the observed value under the null hypothesis. Small p-value represents evidence against H_0 . If the probability is small and below a predefined threshold, then the null hypothesis can be rejected in favor of the alternative hypothesis. Each significance threshold, α , can be translated into a type I error rate, i.e. the probability of a null model being falsely rejected. If the significance level α is set as 5% for each single test, in total, around 5% of the SNPs or voxels will be considered to be associated with the phenotype by chance. For example, in a VBM analysis, if 10K of voxels are tested for significant association and none of them are actually associated, 500 voxels will be incorrectly identified. Such SNPs or voxels are also called false positives (Brett *et al.*, 2003).

To reduce the chance of having too many false positives, a method of setting the significance threshold to control the Family-wise Error Rate (FWER) is needed. FWER is probability that, under the null, one or more tests will be rejected. To control FWER, several methods are available, such as Bonferroni correction. The idea is to replace the significance level α with α/P for each single test, where P is the number of SNPs or voxels tested. However, this method gives very little chance of detecting real effects and it has an assumption that the markers are independent. The assumption is unlikely to be true neither in GWAS nor in VBM. Most GWASs use 5×10^{-8} as a threshold that controls FWER at 5% (Clarke *et al.*, 2011). For VBM studies, the most popular method is random field theory. The idea of RFT is to transform related voxels from the smoothed statistical maps into independent resolution elements or “resels”. A resel is a block of voxels that has the same size as the smoothing kernel full width at half maximum. The smoothed statistical maps with resels can be

thresholded. The Euler characteristic (EC) can be considered the number of blobs after the thresholding. When the threshold is high, the expectation of EC is either zero or one. The expected EC approximately corresponds to the probability of observing a or more blob above a given threshold (FWER probability) (Brett *et al.*, 2003). The expected EC can be calculated using equation (2.4), where R is the number of resels and Z_t is the Z score threshold.

$$E [EC] = R(4\log_e 2)(2\pi)^{-\frac{3}{2}}Z_t e^{-\frac{1}{2}Z_t^2} \quad (2.4)$$

2.4.5 Supervised learning

1) Support vector machine for classification

Linear support vector machine (L-SVM) is a popular supervised classification algorithm as it performs generally well in high-dimensional spaces (Hastie *et al.*, 2009). A classification problem can be either a separable or a non-separable case. In the linearly separable case, the SVM algorithm searches for a decision boundary that maximizes the margin between the two classes. The SVM's margin is a function of a subset of the data points known as Support Vectors SVs. The optimization process is convex. When the two classes are non-separable slack variables are introduced which allow misclassifications to occur. This leads to the introduction of hyper-parameters which dictate how much penalty is occurred in the SVM's objective function if misclassifications occur.

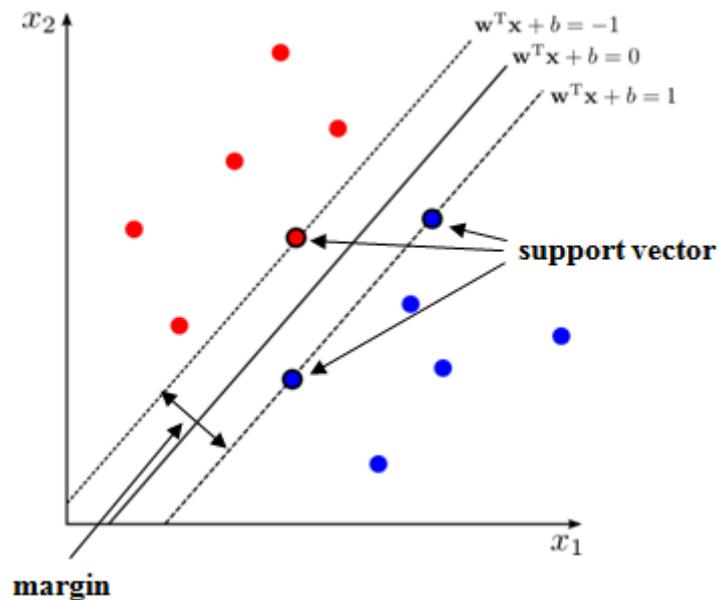


Figure 5 Illustration of support vector machine algorithm

This figure shows a separable case of two classes (+1 in blue and -1 in red). The idea of support vector machine is to search a boundary that maximizes the margin between the two classes. The

decision boundary (the classifier, it will be a hyperplane in high-dimensional spaces) is the solid line, while the broken lines bound the margin that maximized in the algorithm. The points with black circle are support vectors which define the boundary.

2) Statistical Methods for predicting patient survival

For survival problems, Kaplan-Meier (KM) method and Cox’s proportional hazards (PH) regression model are widely used (Fox and Weisberg, 2011; Hastie *et al.*, 2009). In survival studies, the outcome variable is the survival time. Survival time is defined as the time from the starting point to the occurrence of an event of interest. It can be partly observed for some participants (censored data). For example, in a clinical trial, all participants register on the first day and start to receive the same treatment (Figure 6). The clinical trial is supposed to last for 365 days. One participant might experience the event, such as death, after 100 days and the observed survival time of this subject is 100 days. Other participants that survived without having the event occurred until the end of the trail are called “right censored”. Although the exact survival time is unknown, the time that the subject survived is still informative.

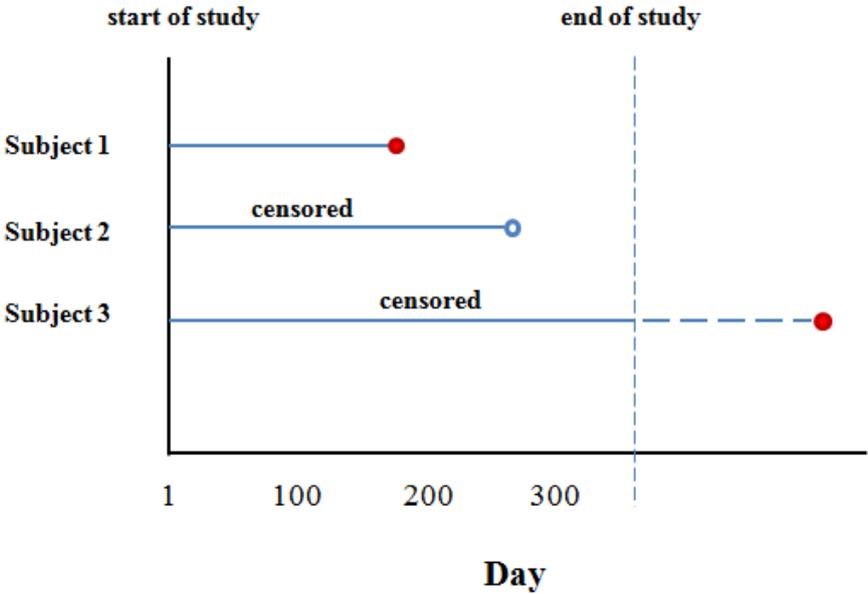


Figure 6 Example of survival data

This figure shows 3 individuals as an example of censored data in survival analysis. Three subjects registered for a clinical trial at the same and received the same treatment. Subject 1 suffered the event (i.e death) before the end of the study. Subject 2 dropped out in the follow-up study. Subject 3 survived until the end of the study and the event occurred afterwards, but his or her exact survival time is unknown. Since subject 2 and subject 3 have no exact survival time, therefore they are the participants whose data are censored.

The Kaplan-Meier method estimates a survival function $S(t)$ (or survival curve). The survival curve represents the probability of surviving beyond a given time t . KM method is useful to compare the outcomes of two groups under different conditions, i.e 2 types of treatment. When a survival analysis examines the relationship between the survival time and the covariates, Cox PH model is preferred. Cox PH model involves hazard function, which measures the instantaneous rate of a subject suffering the event at time t , given that the event has not occurred before. Hazard function is specific to the distributions of survival time, such as the exponential, Weibull and Gompertz distributions (Fox and Weisberg, 2011). If the risk of having the event occurring at a given time t depends on a set of predictors X_1, X_2, \dots, X_p , the hazard function can be written as follows:

$$h(t) = h_0(t) \exp(\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p) \quad (2.5)$$

, where $h(t)$ is the expected hazard at time t , $h_0(t)$ represent the baseline hazard when all predictors value are zero. For two subjects, i and j , the hazard ratio of these two subjects

$$\begin{aligned} \frac{h_i(t)}{h_j(t)} &= \frac{h_0(t) \exp(\beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip})}{h_0(t) \exp(\beta_1 X_{j1} + \beta_2 X_{j2} + \dots + \beta_p X_{jp})} \\ &= \exp\{\beta_1(X_{i1} - X_{j1}) + \beta_2(X_{i2} - X_{j2}) + \dots + \beta_p(X_{ip} - X_{jp})\} \end{aligned} \quad (2.6)$$

is independent of time t . Cox PH model assumes that the effects of covariates, $\beta_1, \beta_2, \dots, \beta_p$ are constant over time and independent of the exact form of $h(t)$. Therefore Cox PH model is a semiparametric model.

2.4.6 Unsupervised algorithm

Popular unsupervised algorithm, such as principal component analysis (PCA) and k-means, have been used in many studies to identify patterns among the input measurements. The advantage of such methods is that they do not rely on the outcomes and therefore could avoid the usage of low accurate clinical diagnosis of neurodegenerative disease.

Event-based model (EBM) is recently developed algorithm and has been used to identify the biomarker changes in Alzheimer's disease progression (Young *et al.*, 2014b). This algorithm learns the sequence of biomarker changes from large cross-sectional data. One event (E) is defined as the switch of one biomarker from normal, which relates to non-pathology, to abnormal state that caused by disease. For each biomarker, the data from all subjects are used to fit two normal distributions correspond to normal and abnormal states. The candidate sequences are generated by Markov chain Monte Carlo algorithm. EBM tests all candidate sequences and identifies a sequence of biomarkers change to maximize data likelihood.

2.5 Hill's causality inference

The identification of the cause of a complex disease is not unified. The Bradford Hill criteria are the most practical and widely accepted guidance for causal inference (Swaen and van Amelsvoort, 2009). In Table 2, each individual criterion will be explained with an example from a smoking and lung cancer study.

	Definition	Example
Temporal Relationship	The causal exposure should precede the caused disease in time.	Most of the lung cancer patients have started smoking before having lung cancer.
Strength	The stronger the association the more likely that the association is causal.	The smoker group has bigger portion cancer patients than non-smoker group.
Dose-response	When a level of exposure increases, risk of disease increases.	There is a positive, linear relationship between the amount cigarette smoked and the incidence of lung cancer.
Consistency	If more studies find similar results, it is more likely that the association is causal.	Previously published studies agree with the conclusion that smoking causes lung cancer. Independent cohorts can replicate the results.
Plausibility	Biological experiments	Evidence of biological experiment that smoking causing tissue damage.
Analogy	Consideration about other potential factors which showing similar effects	Smoking is the most significant predictor for cancer onset after adjustment of other cofounders, like demographic factors.
Coherence	The conclusion does not against current knowledge.	Smoking can cause lung cancer has been long accepted.
Experiment evidence	The elimination of the causal agent can decrease disease rates.	After quitting smoking, the amount of specific toxic products decreases in blood.
Specificity	The exposure associated with a specific outcome.	Lung cancer is a complex disease, not caused by a single agent.

Table 2 Explanation of Hill's criteria with smoking study as an example

3. In-vivo brain neuroimaging provides a gateway for integrating molecular and clinical markers of Alzheimer's disease

Alzheimer's disease (AD) - the most common neurodegenerative disorder is becoming a high socio-economic burden for the Western world. As a response to this challenge, we witness a steadily expanding research in biomarkers aiming at accurate early diagnosis and a reliable prediction of disease progression. The newly proposed diagnostic criteria of the International Working Group for AD describe new phenotypes in AD by integrating novel and established biomarkers with clinical features to further expand the spectrum from the asymptomatic stage to very advanced phase of dementia (Dubois *et al.*, 2014). However, the remaining key challenges are not only to establish the contribution of each biomarker separately, but more importantly to identify how the different biomarkers interact. Understanding the complexity of the biological data and facilitating their interpretation in relationship to the clinical symptoms requires a framework that enables the integration of the existing multiple biomarkers - clinical features, genetic variants, proteins, etc.

In this review, we put forward the hypothesis of using imaging as endophenotype of AD enabling a straightforward interpretation of the high-dimensional genomic and proteomic data (Figure 7). We present the findings from the recent literatures that support this hypothesis and demonstrate bridges between biological data and the clinical manifestation of the disease. In the first section we review papers which illustrated that neuroimaging features are sensitive biomarkers to predict clinical symptoms. In the second and the third sections, we highlight recent studies that go beyond this link and bring evidence for complex interactions between clinical symptoms and biological features, such as particular protein levels and genetic variants. In the last section we explain the current characterization of the process of "healthy" aging to differentiate these from changes associated with abnormal functioning.

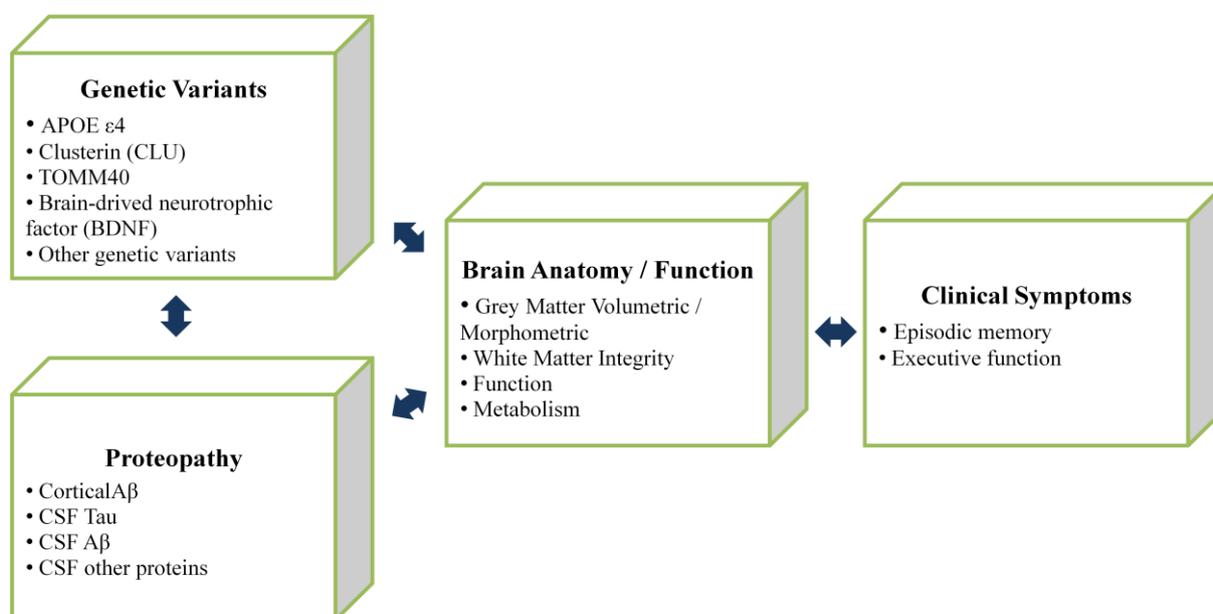


Figure 7 Multiple data modalities in Alzheimer's disease

The relationships between different data modalities have been studied in Alzheimer's disease. These modalities include genetic variants, proteomics, brain anatomy / function and symptoms. Section 3.1 presents the latest findings between brain anatomy / function and symptoms. Section 3.2 presents the associations found between brain anatomy / function and proteomics. Section 3.3 shows the effects of genetic variants on brain anatomy / function. Section 3.4 summarizes aging effects in cognitive normal populations.

3.1 Bridging neuroimaging biomarkers and clinical diagnostic

Recent advances in neuroimaging methods provide an unbiased assessment of brain anatomy that offers a diagnostic window in early stages of AD. Currently, biomarkers are derived from different brain imaging techniques - T1-, diffusion-weighted and functional MRI, Pittsburgh Compound B positron emission tomography (PIB-PET), [18F] 2-fluoro-2-deoxyglucose (FDG) PET. One of the MRI-derived features – the pattern of grey matter (GM) volume loss, has been established as reliable biomarker showing strong correlation with clinical disease staging and neuropathology findings (Braak and Braak, 1995). The atrophy rate of the medial temporal lobe volume is not only closely linked to the individual's cognitive deficits, but has also predictive value for the progression from mild cognitive impairment (MCI) to AD. However the association between neuroimaging biomarkers and clinical diagnosis is far from being complete, this is mainly due to the high level of intersubject variability in both brain changes and clinical measurements. To increase the accuracy of prediction, two approaches have been applied. The first approach is to improve the clinical measurements, for example the inclusion of neuropsychiatric factors such as depression, apathy, anxiety and sleep disturbance (Belleville *et al.*, 2014). The second approach is to introduce novel neuroimaging features,

such as besides assessment of brain anatomy. An example of new neuroimaging feature is to use cerebral blood flow (CBF) based on arterial spin labeling (ASL).

ASL showed a reasonable predictive power to discriminate between dementia patients and healthy controls (Bron *et al.*, 2014). Compared to the diagnostic accuracy of biomarkers using estimates of GM volume, CBF showed a similar performance. An independent analysis, which compared FDG-PET, MRI volumetric variables and PIB-PET for the prediction of conversion from MCI to AD, showed MRI provided the highest predictive accuracy and PIB-PET demonstrated the best sensitivity (Trzepacz *et al.*, 2014).

Although we can improve the prediction accuracy using these two approaches, they are still limited in discovering the biological mechanisms. In the following sections, we focus on studies that bridging biological measurements (genetic and proteomic) and brain anatomy and function under the frame of Alzheimer's disease.

3.2 Bridging protein changes and brain anatomy and function

There is a large body of evidence of the role of misfolded A β 42 plaques and microtubule-associated protein tau neurofibrillary tangles. In vivo, pathological changes can be detected in amyloid burden and CSF protein levels.

The cortical amyloid burden can be directly detected by PET imaging. Currently the most widely used tracer to image cortical A β 42 is ¹¹C-labeled Pittsburgh compound B (PIB) and florbetapir F 18 (also known as [¹⁸F] AV-45). Apart from measuring cortical A β 42 burden using PET imaging directly, cerebrospinal fluid (CSF) A β 42 levels (Strozyk *et al.*, 2003) and phosphorylated tau (pTau) protein at threonine 231 (pTau231) and threonine 181 (pTau181) have been demonstrated to reflect pathologic changes in the brain (Buerger *et al.*, 2006; Tapiola *et al.*, 2009). Although pTau231 protein showed better correlation with neurofibrillary tangle burden at autopsy (Buerger *et al.*, 2007), pTau181 is widely studied so far.

We highlight below the studies that demonstrate the interaction between protein measures and regionally specific changes in brain anatomy and function.

3.2.1 Interaction between Cortical A β 42 peptides and anatomo- functional dysfunction

A β 42 deposition, in healthy old adults, has been shown to modulate the blood-oxygen-level dependent (BOLD) activity in entorhinal regions measured with functional MRI during an episodic memory task. However there was no modulation effect of the BOLD activity in the hippocampus (Huijbers *et al.*, 2014). The authors also showed that that entorhinal regions demonstrating amyloid- β -related dysfunction are directly connected to the neocortical regions of the default network. This result is in

contradiction with another study in an independent cohort across all subjects, including Alzheimer's patients, MCI and healthy elderly (Adriaanse *et al.*, 2014). The latter study found no association between amyloid deposition and the functional connectivity of the default mode network. However, other study support the link between amyloid deposits, tau toxicity (potentiated by A β 42 deposit) and lateral entorhinal cortex dysfunction, see Figure 8 (Khan *et al.*, 2014).

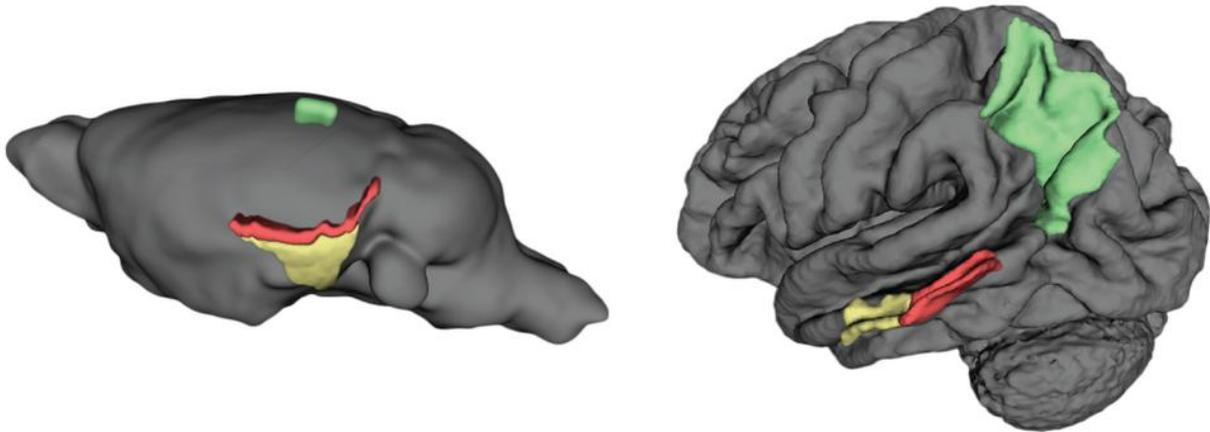


Figure 8 Cerebral blood volume reductions patterns

*To determine whether Tau and APP relate to entorhinal cortex dysfunction, mouse models with mutations in Tau or / and APP genes expressed in the entorhinal cortex were scanned with a high-resolution functional MRI to map metabolic defects. Only mouse model with both mutations showed reliable cerebral blood volume reductions (results shown in rendered three-dimensional volumes, left figure) in lateral entorhinal cortex (yellow), perirhinal cortex (red), posterior parietal cortex (green). These patterns overlap with patterns with patterns observed in preclinical AD. Among the regions affected in the human preclinical group (right figure), dysfunction was observed in entorhinal cortex (yellow), parahippocampal gyrus (red) and precuneus (green). Adapted from (Khan *et al.*, 2014).*

In term of prediction of amyloid status (amyloid-positive or –negative measured by AV45 (florbetapir) PET), basal forebrain cholinergic system volume has been found to be a better anatomical biomarker than hippocampus (Teipel *et al.*, 2014).

3.2.2 Interaction between CSF proteins and anatomo- functional dysfunction

The ratio between CSF phosphorylated tau threonine 181 and A β 42 (pTau181/A β 42) has been proposed as one of the best biomarkers to predict cognitive decline in nondemented older adults (Fagan *et al.*, 2007) and AD conversion risk in MCI group (Buchhave *et al.*, 2012). In a recent study, a significant difference has been found in white matter between MCI low / high ratio groups. The low ratio group showed a significant increase in radial diffusivity in both sides of the corpus callosum and

the superior and inferior longitudinal fasciculus. However, the high-ratio group did not show any difference when compared to a normal group (J.-S. Lim *et al.*, 2014).

CSF tau proteins, including pTau181 and tTau, can be regulated by CSF Apo-E protein, which is the product of apolipoprotein E gene and its level is modulated by apolipoprotein E (APOE) ϵ 4 allele. Baseline CSF Apo-E protein showed a positive association with tTau and pTau181 levels, but not with CSF A β 42. Individuals with low baseline CSF Apo-E levels and high CSF tTau/A β 42 ratio showed higher brain atrophy rate in the fusiform, inferior parietal, inferior temporal, superior frontal, precuneus, middle frontal, and entorhinal areas (Toledo *et al.*, 2014). The effect was independent from APOE- ϵ 4 status.

Similar results have been found with CSF clusterin which interact with CSF A β 42, pTau181 and tTau (total Tau) peptides (Desikan *et al.*, 2014). CSF clusterin predicts brain structure differences between cognitively normal elder participants and MCI in the entorhinal cortex but not in the hippocampus. pTau181 showed an independent additive effect on volume loss in the same brain regions.

3.2.3 Interaction between TDP-43 proteins and anatomo- functional dysfunction

TAR DNA-binding protein of 43 kDa (TDP-43), previously related to frontotemporal dementia has been found to be present in autopsy examinations among pathologically diagnosed AD. A recent study with 342 pathologically diagnosed AD evaluated the effect of TDP-43 on brain structure change. Positive association has been found between TDP-43 burden and medial temporal atrophy and cognitive impairment, see Figure 9 (Desikan *et al.*, 2014). This result suggests a new potential therapeutic target for the treatment of AD. However, more studies are required to understand its mechanism and interaction with other biomarkers.

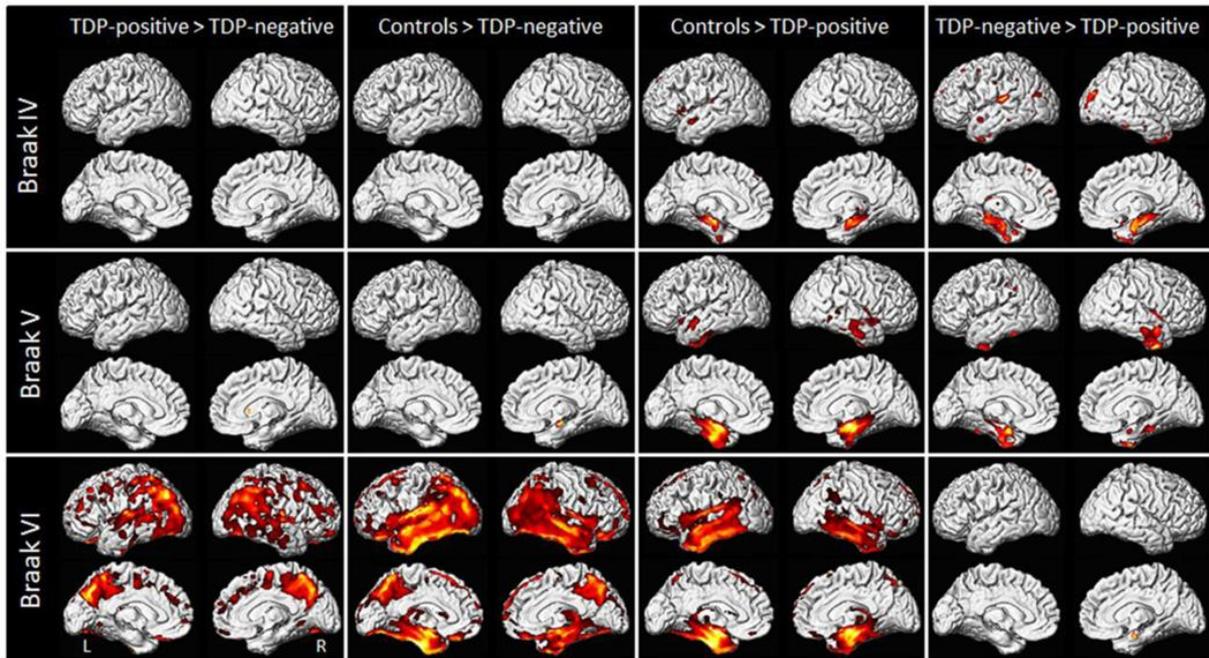


Figure 9 Regional patterns of TDP-associated grey matter volume loss across Braak stages IV to VI. This figure shows atrophy patterns of AD TDP-positive and AD TDP-negative groups diagnosed within Braak stages IV, V and VI, compared with control (pathologically diagnosed between Braak 0-III). Adapted from (Josephs *et al.*, 2014).

3.3 Bridging genetic variants and brain anatomy and function

Up to now, more than 20 common AD genetic risk factors have been found with genome-wide association studies (GWAS). However, apart from APOE- ϵ 4 allele, the effects of these single nucleotide polymorphisms (SNPs) on late onset AD (LOAD) progression have been found to be limited (Carrasquillo *et al.*, 2015). Recent investigations propose to measure the SNPs risk for dementia using neuroimaging features as phenotype by investigating the association between SNPs or SNP-SNP interaction with the brain structure.

3.3.1 Interaction between APOE- ϵ 4 and anatomo-functional dysfunction

The effects of APOE- ϵ 4 on brain structure have been observed not only among LOAD, but also in infant carriers between 2 to 25 months. Risk allele carriers had lower myelin water fraction (MWF) and grey matter volume in areas preferentially affected by AD but greater volume in extensive frontal regions. Infants with ϵ 4 also showed a reduced MWF during development in posterior white matter regions (Dean *et al.*, 2014). Among LOAD “stable” and LOAD “progressors” MCI, ϵ 4 carriers showed a higher hippocampal volume atrophy rate compared to non-carriers, even adjusted for whole brain atrophy rate (Manning *et al.*, 2014; Shi *et al.*, 2014). Among the cross-sectional group (LOAD, MCI and cognitively normal), carriers were found to have significant morphological differences

compared to non-carriers. This result was more significant in homozygotes than heterozygotes (Shi *et al.*, 2014). A voxel based analysis showed a reduction of fractional anisotropy (FA) in the genu of the corpus callosum and the brain stem regions (Newlander *et al.*, 2014). In the TOMM40 gene, the rs10524523 (“523”) variable length poly-T repeat polymorphism has been found to be associated with white matter integrity, regardless of the presence of the APOE- ϵ 4 allele. The main effect of TOMM40 523 has been observed in the right rostral cingulum and the left ventral cingulum FA. The result proved that the short/long repeat length in the genotype of TOMM40 523 was associated with a lower white matter integrity compared with pooled long repeat length groups (Lyll *et al.*, 2014). Clusterin related gene variants (CLU, rs11136000 and rs1532278) have been recently identified for significant combined effects with APOE on both volumetric expansion and lateral ventricle surface morphology in a longitudinal study. However, rs1532278, showed slightly stronger association with ventricular expansion than that of rs11136000 (Roussotte *et al.*, 2014).

3.3.2 Interaction between other common genetic variants and anatomo- functional dysfunction

The genetic variant of one neurotrophic factor (brain-derived neurotrophic factor (BDNF) rs6265), replacing guanine with adenine resulting in a valine (val) to methionine (met) amino acid substitution at codon 66 (val66met), has been proposed to be an indirect moderator of A β 42 neurotoxicity as BDNF is involved in synaptic excitation and neuronal plasticity. Indeed, this hypothesis has been proved among amnesic MCI (aMCI) with high cortical A β 42 load and being met carriers to have greater rate of decline in episodic memory and reduction in hippocampal volume over 36 months (Y. Y. Lim *et al.*, 2014). However, linear met-dose effect was not supported in this analysis of global and local gray and white matter structure. Compared to homozygous groups (val/val and met/met), the val/met group showed an “inverted-U” shaped profile of cortical changes (Forde *et al.*, 2014).

One common variant in the gene encoding the receptors (OPRD1, rs678849), known to be associated with drug addiction and protecting neurons against hypoxic and ischemic stress, has been identified for its association with reduced brain volumes in AD (Roussotte *et al.*, 2014).

SNPs identified with GWAS have a small effect size for explaining clinical changes; therefore more studies are looking at pathway-derived SNP-SNP pairs for investigating genetic effects on brain structure. For example, a gene pair SYNJ2-PI4KA from the inositol phosphate metabolism and the phosphatidylinositol signaling system pathways in KEGG was significantly associated with the rates of change in right and left inferior lateral ventricles volume (Koran, Hohman, Meda, *et al.*, 2014).

3.3.3 Interaction between autosomal dominant Alzheimer disease and anatomo- functional dysfunction

Several studies have been done to study the similarity between autosomal dominant Alzheimer's disease (ADAD) and LOAD of CSF protein levels, amyloid imaging and brain volumetric change. Higher levels of amyloid deposition were detected in basal ganglia in ADAD than in LOAD and increased level of CSF A β 42 appears earlier in ADAD than in LOAD (Fleisher *et al.*, 2012). A recent study compared the two groups in terms of brain functional connectivity and found no differences. For both groups, clinical dementia rating scores correlated negatively with measures of functional connectivity (Thomas *et al.*, 2014).

3.3.4 Genetic variants regulate cortical amyloid burden

Cortical amyloid burden can be measured with PIB-PET and is considered as an endophenotype to evaluate the effects of polymorphism. APOE- ϵ 4 has been found to have positive association with the cortical A β 42 burden level among healthy controls and MCI (Drzezga *et al.*, 2009; Rowe *et al.*, 2010). But this genotype effect on the probable AD group was not consistent between studies (Drzezga *et al.*, 2009; Rowe *et al.*, 2010). By studying a population of probable AD, comparing APOE- ϵ 4 non carriers with ϵ 4 carriers, significantly higher levels of A β 42 burden were found in the right lateral frontotemporal regions and greater hypometabolism in the cortical areas (Lehmann *et al.*, 2014).

One polymorphism upstream of butyrylcholinesterase (BCHE) on chromosome 3 has been found to have a novel modulating effect with cortical amyloid load, independently from APOE- ϵ 4 genotype (Ramanan *et al.*, 2014). A polymorphism in GSK-3 (glycogen synthase kinase 3), implicated in both Tau and amyloid pathology, was found to be involved in 3 pairs of SNP-SNP interaction with baseline cortical amyloid deposition, independently from APOE genotype (Hohman *et al.*, 2014). Similar SNP-SNP interaction in amyloidogenesis has revealed the effect of the calcium channels pathway on amyloid deposition (Koran, Hohman and Thornton-Wells, 2014).

3.4 Aging related biomarker changes

Several studies have been performed to characterize normal aging effects on the human brain, especially on task related activation, age-related volumetric and morphometric changes to separate effects due to the disease from normal ageing. Most of these studies show a significant effect of normal ageing on the same regions (e.g. hippocampus, entorhinal cortex) that were highlighted previously as being involved in AD.

Cognitively normal older adults showed significant lower hippocampal volume and thinner entorhinal cortices compared to young adults and reduced activation during an episodic memory task (Huijbers *et al.*, 2014).

The age-related structural change was also observed in a longitudinal study performed over 2 years. Compared to other regions, including hippocampus, bilateral transverse temporal regions showed the

fastest atrophy, with a decrease in thickness of 4.6% and 4.9% per year in the left and right hemispheres. Regions like the entorhinal cortices, the right hippocampus and the right precentral area showed a age-related change in the rate of atrophy (Jiang *et al.*, 2014). Sex-related cortical atrophy rate differences have also been identified among these non-demented elder people, see Figure 10 (Jiang *et al.*, 2014).

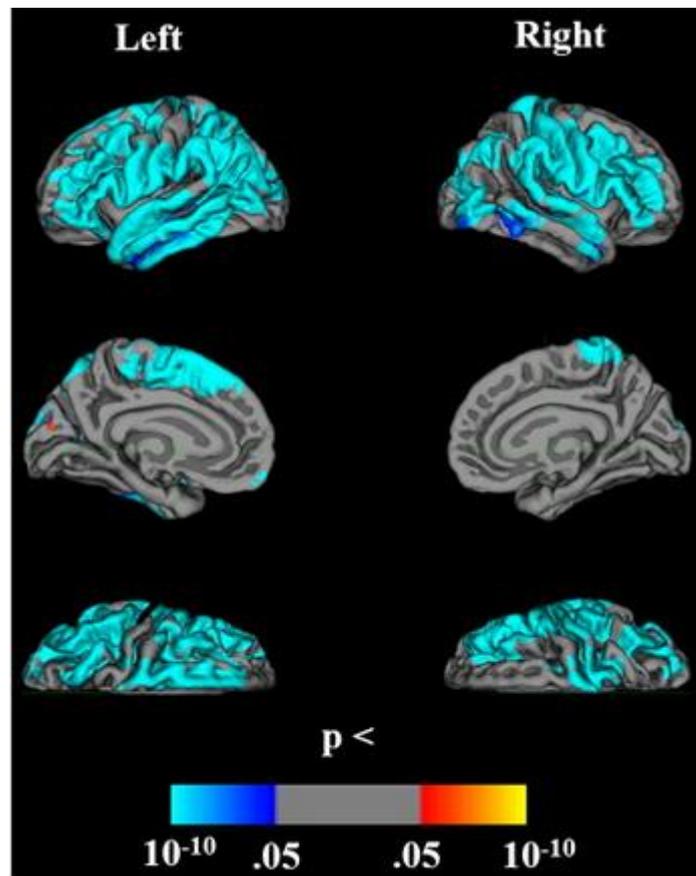


Figure 10 Cortical atrophy over 2 years in non-demented elderly adults

Cortical thickness between baseline and follow-up has been found based on non-demented aging population. The most significant change was found in transverse temporal region. Adapted from (Jiang et al., 2014).

A separate longitudinal study, over 4 years, among middle-aged healthy participants has been done to study the normal aging effect on white matter. Significant shrinkage in white matter volume has been found in frontal, temporal, and cerebella regions (Ly *et al.*, 2014).

3.5 Conclusion

There is no doubt that neuroimaging based biomarkers are useful endophenotypes that can bring together information from other modalities and bridge the gaps between different scales of

observations in the brain. However, neuroimaging biomarkers still suffer several limitations, which are mainly due to the high degree of variability between multi-centers (Marchewka *et al.*, 2014), cohorts and individuals. New techniques based on tissue properties and quantitative MRI (e.g., myelin, iron, and water content) are now providing compelling alternatives (Draganski *et al.*, 2014). By combining different biomarkers, we can observe apparent discrepancies which can be seen at first as a lack of accuracy but it can also reflect the existence of different subtypes of Alzheimer's disease (Young *et al.*, 2014a). The identification of pathological subtypes which are not clinical subtypes is a hard challenge. But methods that decompose the individual differences to create patients stratification can be applied (Kherif *et al.*, 2009). One example is the evaluation of APOE genotype's impact on A β 42 burden between subtypes of pathological defined Alzheimer's disease may reveal the true APOE genotype effect on Alzheimer's disease.

4. Automated detections of asymptomatic individuals at risk for dementia due to Alzheimer's disease

4.1 Motivation

Symptoms can appear decades after the biological change (Dubois *et al.*, 2016). To provide early treatment and therapy to patients to stop the disease's progression requires a diagnostic tool to identify asymptomatic individuals. The issues with applying research criteria in clinical practice are twofold: firstly, the pathology-related biomarkers are not easily obtainable: collecting CSF samples are invasive and amyloid PET is expensive. To overcome these issues, easy accessible biomarkers, such as blood-based biomarkers or pathology-related structural MRI markers that are able to identify subjects at preclinical stage of AD, are urgently required. Secondly, the mechanism of pathological biomarkers' association with neuronal loss and memory decline is still unclear. Therefore, specific designs of clinical trials with Intervention outcomes still need to be fully studied.

To identify cognitive normal people who are at risk of developing MCI or AD, several studies (Jack *et al.*, 2014) used regional volume obtained from MRI and set a threshold to split subjects at risk (Coupé *et al.*, 2015; Jack *et al.*, 2014; Stephan *et al.*, 2015), such as entorhinal cortex and hippocampus volume. However, regional volumes in hippocampus or entorhinal cortex are not specific to AD as other neurodegenerative diseases, like tauopathy, can cause volumetric change in these regions. Other machine learning methods, like classification based on a clinical diagnosed training set, have been proposed as well to predict subjects who are at risk (Davatzikos *et al.*, 2009; Vemuri *et al.*, 2009). The main drawback of this method is that the training step still relies on clinical diagnosis, which only appears decades after the biological change. Therefore, the circular argument remains. Recent studies have focused on subjects who have biological change, like A β 42 and Tau proteins, and have investigated the transition pathways from cognitive normal to clinical appearance (Jack *et al.*, 2014). As mentioned previously, this method is invasive and expensive, therefore being difficult to apply in the preclinical stage among general populations.

4.2 Introduction

Current criteria for clinical diagnosis of Alzheimer's disease (AD) lack the power to reveal pathological brain changes that identify asymptomatic individuals at risk, before the stage at which symptoms become manifest and progress first to mild cognitive impairment (MCI) then dementia (Dubois *et al.*, 2014). AD pathology, like that of Parkinson's disease, precedes up to 20 years before the onset of cognitive symptoms (Dubois *et al.*, 2016), testifying again to the significant redundancy of brain organization with a resulting capacity for reorganization in the face of pathology (Jack *et al.*, 2013). In the light of this fact, it is preferable to propose a diagnostic strategy that uses pathological

information. A unified conception as evidence of A β 42 or Tau pathology from cerebrospinal fluid (CSF) or molecular neuroimaging has been proposed in defining individuals who are ASymptomatic but at Risk (ASR) for AD in the latest version of preclinical AD research criteria in 2016 (Dubois *et al.*, 2016). In parallel, there is a growing interest in using topographical markers (TM) in terms of atrophy pattern derived from structural MRI, which has the advantages of being non-invasive and sensitive to early changes in AD (Coupé *et al.*, 2015).

This study proposes a novel classifier that uses TM from structural data to predict AD pathology. TM alone has been criticised for being insufficient to identify the presence of preclinical AD (Dubois *et al.*, 2016). To improve the usage of TM in preclinical stage, we need to go beyond regional measures (Stephan *et al.*, 2015) and enhance TM to predict pathology (TMPP). In contrast to other studies in which the classifiers were built based on symptoms only (Davatzikos *et al.*, 2009; Vemuri *et al.*, 2009), we construct a classifier trained using clinically and pathologically confirmed cases (Klöppel *et al.*, 2008) to predict pathology. The use of this class of individuals with both diagnostic criteria is novel because, as pointed out by Sperling *et al.* (2011), AD pathophysiology is “necessary but not sufficient to produce the clinical manifestations of AD” (Sperling *et al.*, 2011). Our approach circumvents the circular argument that is implicit in identifying pathological, hence diagnostic and prognostic biomarkers of a disease by correlation with clinical diagnosis alone.

The classifier was used to stratify new study participants from two independent data sets, categorized clinically as either cognitively normal (CN) or AD. We then make a direct comparison between clinical diagnoses and the classifier predictions (pathology label indicated by “P+” or “P-,” respectively). Such a comparison could result in up to 4 groupings: those individuals with matched diagnoses between clinical and prediction (CN_P- cognitively normal predicted negative pathology and AD_P+ clinically probable AD predicted positive pathology) and those in whom there is a mismatch (CN_P+ cognitively normal with a predicted positive pathology and AD_P- clinically probable AD with predicted negative pathology).

In this study, we sought to test whether the predicted AD pathology of CN individuals (CN_P+ and CN_P-) correlates to having risk to develop MCI or clinical AD later in life. Our hypothesis was that the CN_P+ group has a higher risk and proportion of clinical conversion than CN_P-. We also examined risk factors such as apolipoprotein E ϵ 4 (APOE- ϵ 4) genotype (Roses, 2006) and other single nucleotide polymorphisms (SNPs) (Carrasquillo *et al.*, 2015; Slegers *et al.*, 2015) associated with AD, to identify additional prognostic factors for clinical conversion. In addition, we expected the patterns of neuronal loss of CN_P+ individuals at baseline are closer to AD_P+ and they had a progressive change over time compare with CN_P-. To further characterize the predicted groups, we reported AD-related biomarkers levels, glucose metabolic reductions and cognitive performance.

4.3 Method

4.3.1 Data and Materials

We used three independent cohorts: i) pathologically proven by post-mortem examination (PPG), ii) Alzheimer's Disease Neuroimaging Initiative (ADNI) data and iii) the French population-based Three-City (3C), Dijon cohort.(Glymour *et al.*, 2012; Klöppel *et al.*, 2008)

1) Pathologically proven group (PPG)

The group with pathologically confirmed diagnoses comprised 18 AD patients and 15 age and gender matched CN controls.(Klöppel *et al.*, 2008) Cognitive normal were examined either by no evidence of cognitive decline on extended clinical follow-up or via autopsy confirmation, when available. T1 weighted MR images were acquired on three different 1.5 Tesla scanners (Table 3) on average 3.5 years (mean) prior to death. Details are described in Klöppel et al (group 2 in the study, Dementia Research Centre, University College London).(Klöppel *et al.*, 2008)

	Pathologically Proven group		ADNI Cohort		3C Cohort
	CN (N=15)	AD (N=18)	Clinical CN (N=356)	Clinical AD (N=283)	Clinical CN (N=1448)
Age – yr	64.4±11.1	65.9±8.7	75.1± 5.7	75.0 ± 7.6	72.3 ± 4.1
Male sex – no. (%)	9 (60.0)	12 (66.6)	179 (50.3)	151 (53.3)	560 (38.7)
Baseline MMSE score ^b			29.1 ± 1.1	23.2 ± 2.1	28.2 ± 1.2
Education – yr			16.2 ± 2.7	15.1 ± 3.0	
APOE-ε4 carrier			95 (27.0)	175 (67.8)	319 (22.2)
APOE-ε4 1 copy			84 (23.9)	123 (47.7)	
APOE-ε4 2 copy			11 (3.1)	52 (20.2)	

Table 3 PPG, ADNI and 3C cohorts baseline demographic information

^a Plus-minus values are means ± SD.

^b Scores on the Mini-Mental State Examination (MMSE) range from 1 to 30, with higher scores indicating greater cognitive function.

2) Alzheimer's Disease Neuroimaging Initiative

Data used in the preparation of this article were obtained from the ADNI database (www.adni.loni.usc.edu). 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 resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure progression of MCI and early AD.

We used T1-weighted MRI from clinically diagnosed subjects, 283 clinical AD and 356 CN, downloaded on 1 August 2013. Diagnoses in these individuals were based on the mini-mental state examination (MMSE), clinical dementia rating scale (CDR) and NINCDS/ADRDA criteria.(Mckhann *et al.*, 1984) Exclusion criteria are defined in the ADNI procedures manual(‘ADNI procedures manual’, n.d.). Demographic information is listed in Table 3 for case and control groups, diagnosed at baseline. There is no difference between the two groups in age or gender but there is a difference in years of education ($P < 0.001$ by Wilcoxon rank-sum test).

ADNI T1 weighted MR images were acquired at multiple occasions. The first scan of each subject was used as baseline. The final MR scans were collected mean 3.1 years after baseline. Genotyping was performed on 499 subjects. CSF examinations included assays of A β 42, total Tau, and phosphorylated Tau (pTau) proteins at baseline (Shaw *et al.*, 2009). Cortical A β 42 burden was measured by florbetapir F 18 (also known as [18F] AV45) (‘ADNI procedures manual’, n.d.). The level of A β 42 burden is the average AV45 standard uptake value ratio of frontal, anterior cingulate, precuneus, and parietal cortex relative to the cerebellum at baseline. The severity of hypometabolism from [18F]2-fluoro-2-deoxyglucose (FDG) PET was the average FDG-PET of angular, temporal, and posterior cingulate at baseline (‘ADNI procedures manual’, n.d.). MMSE and other clinical scores were downloaded from the ADNI website at baseline and for the following 4 years.

3) Three-City

We used the first recorded MRI from the 1463 aged CN from Dijon cohort of the French 3C study, who were clinically diagnosed at baseline (Table 3) (Alpérovitch *et al.*, 2002). In contrast to ADNI, the 3C study does not include any criteria for MCI diagnosis. Therefore, subjects are considered as cognitively normal until conversion to dementia. We excluded 15 subjects who developed symptoms related to non-AD types of dementia in the follow up study. Study details are described in Stephan *et al* (Stephan *et al.*, 2015).

4.3.2 Data Processing

1) Image processing

Grey matter volumes were calculated in all study individuals using SPM12, an open source software package written in Matlab (<http://www.fil.ion.ucl.ac.uk/spm/software/spm12/>). Neuroimaging processing steps are explained in section 2.4.3 with more details.

2) Genetic variant processing

Single nucleotide polymorphisms (SNP) were typed for ADNI subjects and all samples underwent quality control using PLINK software, passing the following criteria: call rate higher than 90%, inbreeding coefficient (≤ 0.2), concordant sex. (Purcell *et al.*, 2007)

APOE- $\epsilon 4$ genotype status, defined by rs429358 and rs7412 were measured (more details in Supplementary material). In addition, we selected 20 AD associated SNPs, in addition to APOE, from previous studies (Harold *et al.*, 2009; Hollingworth *et al.*, 2011; Lambert *et al.*, 2009, 2013; Naj *et al.*, 2011; Seshadri *et al.*, 2010). In total, 16 high quality, well-tagged SNPs were used in further analysis (see Table 6 and the Supplementary material for more details).

4.3.3 Statistical Analysis

1) Classifier training

To avoid cohort bias, before training a classifier to predict pathology, we first applied two-sample T-test and identified regions with different distributions in grey matter volume between the PPG and ADNI cohorts, adjusting for age, gender and total intracranial volume (TIV). We excluded these regions in the training and prediction process. In total, 492448 voxels from grey matter were used. We conducted within-subject standardization with mean zero and standard deviation one in all individuals from the three cohorts. Standardized images were adjusted for the effects of age and total intracranial volume (TIV) using pathologically verified cognitively normal controls with a linear model.

A library for support vector machines (SVM), libsvm (Chang and Lin, 2011), was used for training the classifier. We trained a linear SVM on the scans from the PPG. One parameter of the classifier, cost C , was selected based on the estimated performance, accuracy, of the models by a line search, using leave-one-out cross-validation.

2) Classifier prediction

The constructed classifier was then used on ADNI and 3C cohorts in order to predict the presence of AD associated pathology of CN individuals at baseline. Clinically diagnosed CN and AD were stratified based on predicted AD pathology into subgroups.

3) Survival analysis

CN individuals at baseline were stratified based on predicted AD pathology into two groups: CN with positive (CN_P+) and negative group (CN_P-). The event-free survival curves of these two groups in the follow-up were estimated by Kaplan-Meier method. Event-free survival is defined as the time from baseline until the first clinical diagnosis of MCI, when available, or AD. Log-rank tests were used to compare the survival curves for the two groups. We used Cox's proportional hazards regression to

estimate the relative risk of developing MCI or AD-type dementia between CN_P+ and CN_P- groups. The relative risk was calculated after adjusting for several disease-related confounders, including age, gender, APOE- ϵ 4 genotype and years of education. In addition to APOE- ϵ 4, other AD-related common genetic risk variants (16 SNPs) were tested for their association with clinical conversion to MCI or dementia due to AD, separately and in combination by calculating a weighted genetic risk score (wGRS, more details are explained in section 2.4.2) (Carrasquillo *et al.*, 2015). Kaplan-Meier curves and Cox's proportional hazards regression were constructed with the R package *survival* (Therneau and Lumley, 2016).

4) Group level grey matter volume difference

SVM classifier predicted AD pathology for CN on individual basis. As subjects may be classified due to different atrophy patterns, to reveal a common atrophy pattern at group level, we compared subgroups (AD_P+, AD_P-, CN_P+ and CN_P-), using mass-univariate voxel-based morphometry (VBM) to investigate topographical, grey matter volume, difference at voxel level. Between-group analyses were done in SPM12.

We hypothesised that CN_P+ atrophied in regions similar to that of PPG AD group compared with CN_P- , affected preferentially at early stage of AD. Affected regions then expanded during the disease progression. To test these two hypotheses, we characterized atrophy patterns at baseline and follow-up. A reference atrophy pattern generated using the PPG was adjusted for age, gender and total intracranial volume (TIV). For comparisons between the subgroups, age, gender, TIV and magnetic field strength were used as covariates, except for the 3C, Dijon cohort as in that a 1.5-T Magnetom was used exclusively.

5) Biomarkers, glucose metabolic levels and cognitive performance

We assessed the differences between predicted groups: in cortical A β 42 burden from AV45; in glucose metabolism levels from FDG-PET; in CSF measures of A β 42, pTau and total tau. We hypothesized that AD_P- and CN_P+ groups show intermediate values between AD_P+ and CN_P- groups. To assess cognitive differences between predicted subgroups, we used total score of each clinical test measured at baseline and a follow up study. For all the comparisons between subgroups, Wilcoxon rank-sum test was applied.

4.4 Results

4.4.1 Classifier training

We used all 33 PPG subjects to train and validate a SVM classifier to discriminate between AD and CN. We achieved 87% accuracy using leave one out cross-validation (cost=1e-4, sensitivity=88.8%, specificity=86.6%).

4.4.2 Classifier prediction

The classifier was then used on each ADNI and 3C subject to predict pathology based on anatomical patterns of atrophy. ADNI subjects fell into 4 groups: 313 CN_P-; 180 AD_P+; 103 clinically diagnosed AD subjects were classified as AD pathology negative (AD_P-); 43 clinically diagnosed CN subjects with an AD specific atrophy pattern who were judged to be pathologically positive (CN_P+). The demographic information is shown in Table 4. AD_P+ group is younger than AD_P- groups (P=0.002 by Wilcoxon rank-sum test). This difference in age was not observed between CN_P+ and CN_P-. Among 1448 CN subjects from the 3C cohort, we predicted 523 subjects to be CN_P+. The CN_P- group is significantly younger than the CN_P+ group, indicating a mild age effect in the classifier (P=0.004 by Wilcoxon rank-sum test). This observation suggests a non-linear method can be applied when the training sample size is big enough.

4.4.3 Survival analysis

We assessed whether the classifier prediction was associated with the development of clinical symptoms by comparing the event-free survival rates of CN_P- and CN_P+ subjects, where the event was defined as clinical conversion to MCI, when available, or AD. The event-free survival rate was compared after five years in ADNI and ten years in 3C, due to the difference in mean follow-up time in the two cohorts (3.7 and 7.2 years respectively). In ADNI, the five-year event-free survival rate was higher in CN_P- (82% [95% CI, 76-88%]) than CN_P+ (63% [95% CI, 43-92%]). Similar results were found in the 3C data, where the ten-year event-free survival of CN_P- was 95% [95% CI, 93-97%]; while in CN_P+ it was 62% [95% CI, 42-93%]. The overall event-free survival rate was significantly higher in CN_P- than in CN_P+ subjects in both cohorts (Figure 11, P<0.001 by Log-rank test). This difference remained significant when considering only conversion to MCI in ADNI (P<0.001 by Log-rank test, MCI diagnosis was unavailable in 3C data).

Multivariate analyses using Cox proportional hazards models (adjusted for age, gender, education and genotypes) showed that CN_P+ subjects had a significantly higher risk of developing MCI or clinical AD than those who were CN_P-, 3.16 times higher risk in ADNI (MCI or clinical AD conversion) and 2.37 times in the 3C study (dementia conversion, Table 5, model 1, 3). APOE-ε4 showed significant impact on conversion with an effect size equal to 2.27 using 3C data ([95% CI, 1.40-3.70] P=0.001 by Wald test, Table 5, model 3). This effect was only observed in APOE-ε4 2 copy carriers in ADNI data (P<0.001 by Wald test, Table 5, model 1). We noticed a trend that the genetic risk score, wGRS based on 16 AD-related SNPs, had a positive association with conversion from CN to MCI or clinical AD

but not significantly ($P=0.08$ by Wald test, Table 5, model 2). Further, each SNP (apart from APOE- $\epsilon 4$) was examined individually but none showed prognostic power for conversion (uncorrected $P>0.05$ by Wald test, Table 6). Among AD-related confounders, age and education showed significant association with clinical AD conversion in 3C ($P<0.001$ and $P=0.03$ for age and education respectively, by Wald test).

4.4.4 Group level grey matter volume difference

Comparison between the pathologically validated AD and CN groups (Figure 12 A left) revealed atrophy restricted to amygdala, parahippocampal gyrus, superior, middle and inferior temporal gyri, inferior and middle occipital lobes, thalamus and median cingulate regions. The atrophy pattern of clinically diagnosed AD compared with CN groups from ADNI study occurred mainly in hippocampus, amygdala, entorhinal cortex and parahippocampal gyrus (Figure 12 A right).

Compared with CN_P-, AD_P+ showed significant atrophy in both mesial temporal lobe regions (Figure 12 B left). Only the hippocampal regions showed significant difference when comparing AD_P+ with CN_P+ (Figure 12 B right).

CN_P+ group when compared with CN_P- was characterised by significantly lower grey matter volume in both mesial temporal lobe regions comprising the fusiform gyri and the entorhinal cortex as well as the inferior, middle and superior temporal gyri that nevertheless was less marked than in AD_P+ (Figure 12 C left). A difference in atrophy over time between CN_P+ and CN_P- was identified using subjects' latest MR scans. Atrophy in the CN_P+ group extended over time into hippocampi and mesial temporal areas in both hemispheres next to the anterior insula and thalamus (Figure 12 C right, the MR scans of CN_P+ and CN_P- mean (\pm SD) durations are 2.3 ± 1.5 and 3.2 ± 2.1).

In the 3C cohort, grey matter volume differences in the CN_P+ compared with CN_P- (Figure 13 A) were mostly observed in both temporal lobes especially (hippocampi, entorhinal cortices, parahippocampal gyri, fusiform gyri) and also extended to limbic areas (insulae, anterior and posterior cingulate cortices) as well as ventromedial and orbitofrontal cortices.

The AD_P- group from the ADNI cohort showed less atrophy predominantly in both lateral temporal lobes, excluding the hippocampal regions, when compared with AD_P+. Additionally, higher volumes were estimated in limbic regions including the posterior insula, posterior cingulate cortex abutting the parietal cortices and in both thalami. However, the AD_P- group showed significant atrophy in both mesial temporal lobe regions including the hippocampus, amygdala, entorhinal cortex and parahippocampal gyrus compared with CN_P- (Figure 13 B, C).

Clinical diagnostic Prediction	ADNI Cohort				3C Cohort	
	Clinical AD (n=283)		Clinical CN (n=356)		Clinical CN (n=1448)	
	P+ (n=180)	P- (n=103)	P+ (n=43)	P- (n=313)	P+ (n=523)	P- (n=925)
Age – yr	74.0 ± 7.6	76.9 ± 7.3	75.2 ± 5.8	75.1 ± 5.7	72.7 ± 4.1	72.0 ± 4.0
Male sex – no. (%)	118 (65.5)	33 (32.0)	29 (67.4)	150 (47.9)	315 (60.2)	245 (26.5)
Baseline MMSE score ^b	23.1 ± 2.1	23.4 ± 2.1	29.1 ± 1.0	29.1 ± 1.1	28.1 ± 1.2	28.2 ± 1.2
APOE-ε4 – no. (%)	110 (68.3)	65 (67.0)	13 (31.0)	82 (26.5)	122 (23.4)	197 (21.5)
# conversion			11 (25.6)	47 (15)	40 (7.6)	32 (3.4)
AV45 ^c	1.4 ± 0.22 (n=71)	1.35 ± 0.2 (n=24)	1.11 ± 0.18 (n=25)	1.12 ± 0.2 (n=102)		
FDG ^d	5.2 ± 0.67 (n=135)	5.77 ± 0.63 (n=57)	6.35 ± 0.6 (n=34)	6.54 ± 0.54 (n=198)		
CSF tau ^e	110.53±48.3	137.15±66.8	87±25.98	68.92±30.6		
CSF pTau ^e	41.54±20.2	42.07±19.8	27±6.8	24.83±14.9		
CSF Aβ42 ^e	136.71±34.9	151.83±47.8	211.4±66.5	205.9±54.7		

Table 4 Predicted groups of ADNI and 3C cohorts baseline demographic information ^a

^a Plus-minus values are means ± SD.

^b Scores on the Mini-Mental State Examination (MMSE) range from 1 to 30, with higher scores indicating greater cognitive function.

^c Average AV45 SUVR of frontal, anterior cingulate, precuneus, and parietal cortex relative to the cerebellum at baseline

^d Average FDG-PET of angular, temporal, and posterior cingulate at baseline

^e The unit of CSF tau, Aβ42 and pTau is pg/mL; subjects the number of subjects in each group is 59, 41, 5, 108 (ordered as AD_P+, AD_P-, CN_P+ and CN_P-)

Variable	ADNI cohort				3C cohort	
	Model 1		Model 2		Model 3	
	Hazard Ratio (95% CI)	P Value	Hazard Ratio (95% CI)	P Value	Hazard Ratio (95% CI)	P Value
Clinical CN, Classified P+ (CN_P+)	3.16 (1.58-6.32)	0.001	3.21 (1.50-6.88)	0.003	2.37 (1.45-3.89)	<0.001
Female	0.6 (0.35-1.06)	0.07	0.54 (0.29-1.0)	0.05	1.08 (0.64-1.81)	0.78
Education, per 1-yr increase	0.92 (0.84-1.02)	0.10	0.92 (0.83-1.01)	0.1	0.84 (0.71-0.98)	0.03
Age, per 1-yr increase	1.03 (0.98-1.09)	0.19	1.07 (1.01-1.13)	0.02	1.17 (1.10-1.24)	<0.001
APOE-ε4 dose						
2	7.09 (2.46-20.47)	<0.001	8.73 (2.53-30.05)	<0.001		
1	1.51 (0.85-2.68)	0.15	1.41 (0.76-2.66)	0.28		
1 or 2^a					2.27 (1.40-3.70)	0.001
Genetic risk score (16 SNPs)	-	-	2.21 (0.92-5.30)	0.08		

Table 5 Effect of risk factors on conversion from cognitive normal to MCI/AD in ADNI and 3C cohorts

^a 3C cohort has no information about the number of APOE-ε4 allele

<i>Gene & SNP & Risk Allele & OR *</i>				<i>Hazard Ratio (95% CI)</i>	<i>P Value</i>
APOE	rs429358 rs7412	ε4	2.53		
CLU	rs11136000	G	1.22	0.97 (0.64-1.48)	0.90
PICALM	rs3851179	G	1.25	1.33 (0.86-2.06)	0.20
CR1	rs3818361	A	1.19	1.33 (0.74-2.40)	0.33
BIN1	rs744373	G	1.15	0.97 (0.64-1.47)	0.90
ABCA7	rs3764650	G	1.22	0.96 (0.49-1.88)	0.89
MS4A6A	rs610932	C	1.15	1.14 (0.74-1.76)	0.56
EPHA1	rs11767557	A	1.15	1.60 (0.97-2.64)	0.07
CD33	rs3865444	C	1.09	1.30 (0.83-2.04)	0.25
SLC24A4-RIN3	rs10498633	G	1.07	1.32 (0.77-2.24)	0.31
FERMT2	rs17125944	C	1.17	0.91 (0.38-2.18)	0.84
CD2AP	rs9349407	C	1.07	0.99 (0.61-1.61)	0.97
PTK2B	rs28834970	C	1.11	0.90 (0.60-1.35)	0.61
MEF2C	rs190982	A	1.07	0.74 (0.47-1.16)	0.19
ZCWPW1	rs1476679	T	1.12	1.52 (0.96-2.42)	0.07
CELF1	rs10838725	C	1.09	1.01 (0.66-1.54)	0.98
CASS4	rs7274581	T	1.12	1.36 (0.61-3.00)	0.45

Table 6 Effect of each SNP on conversion from cognitive normal to MCI/AD in ADNI cohort

Hazard ratio (HR) and 95% CI were estimated for the genetic variants.

** Hazard ratios correspond to and additional allele and the full models were adjusted for predicted groups, gender, age, and years of education.*

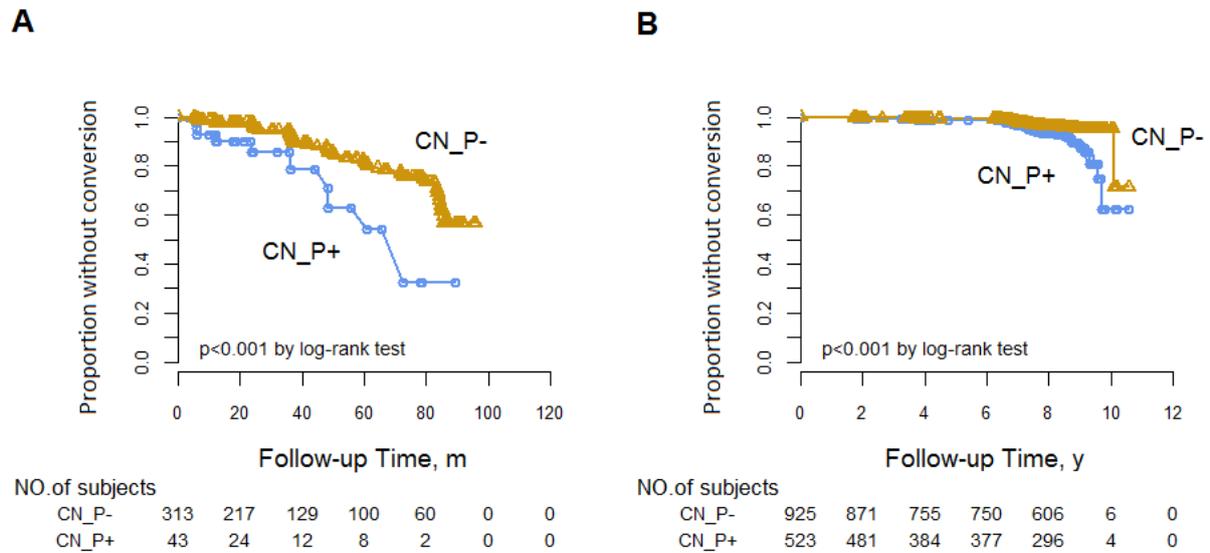


Figure 11 KM survival curves of predicted CN groups to MCI or AD in ADNI and 3C cohorts

A Kaplan-Meier plot with the time after baseline at conversion to MCI or clinical AD for cognitive normal individuals at baseline prediction who were predicted to have pathology (CN_P+, $n = 43$ in ADNI and $n = 523$ in 3C) and those predicted to have non-pathology (CN_P-, $n = 313$ in ADNI and $n = 925$ in 3C). Panel A and B show Kaplan-Meier curves in ADNI cohort and 3C cohort respectively.

4.4.5 Biomarkers, glucose metabolic levels and cognitive performance

A β 42 burden measured from AV45 were higher in AD subjects with no difference between P- and P+. Glucose metabolism levels from FDG-PET were lower in AD subjects, AD_P+ having a significantly lower level than AD_P-. CSF levels of both total tau and pTau were significantly higher in both AD_P- and AD_P+ than in CN_P-, with intermediate values in CN_P+. Although pTau levels were similar in AD_P- and AD_P+, total tau was significantly higher in AD_P- than AD_P+. CSF levels of A β 42 were lower in AD than in CN, with no significant difference between classifier-predicted groups. Mean values are shown in Table 4, for the descriptive statistics see Figure 14.

Total MMSE scores collected from baseline till 48 months later were compared between each of the four groups (Figure 15). The AD_P+ group showed the lowest values throughout follow up. The AD_P- group showed intermediate values between AD_P+ and the two clinically CN groups. CN_P+ individuals had values significantly lower than CN_P- at 48 months. This difference in MMSE is evidence which supports prognostic value of our stratification. Other clinical scores were compared between predicted groups (Supplementary Figure 2).

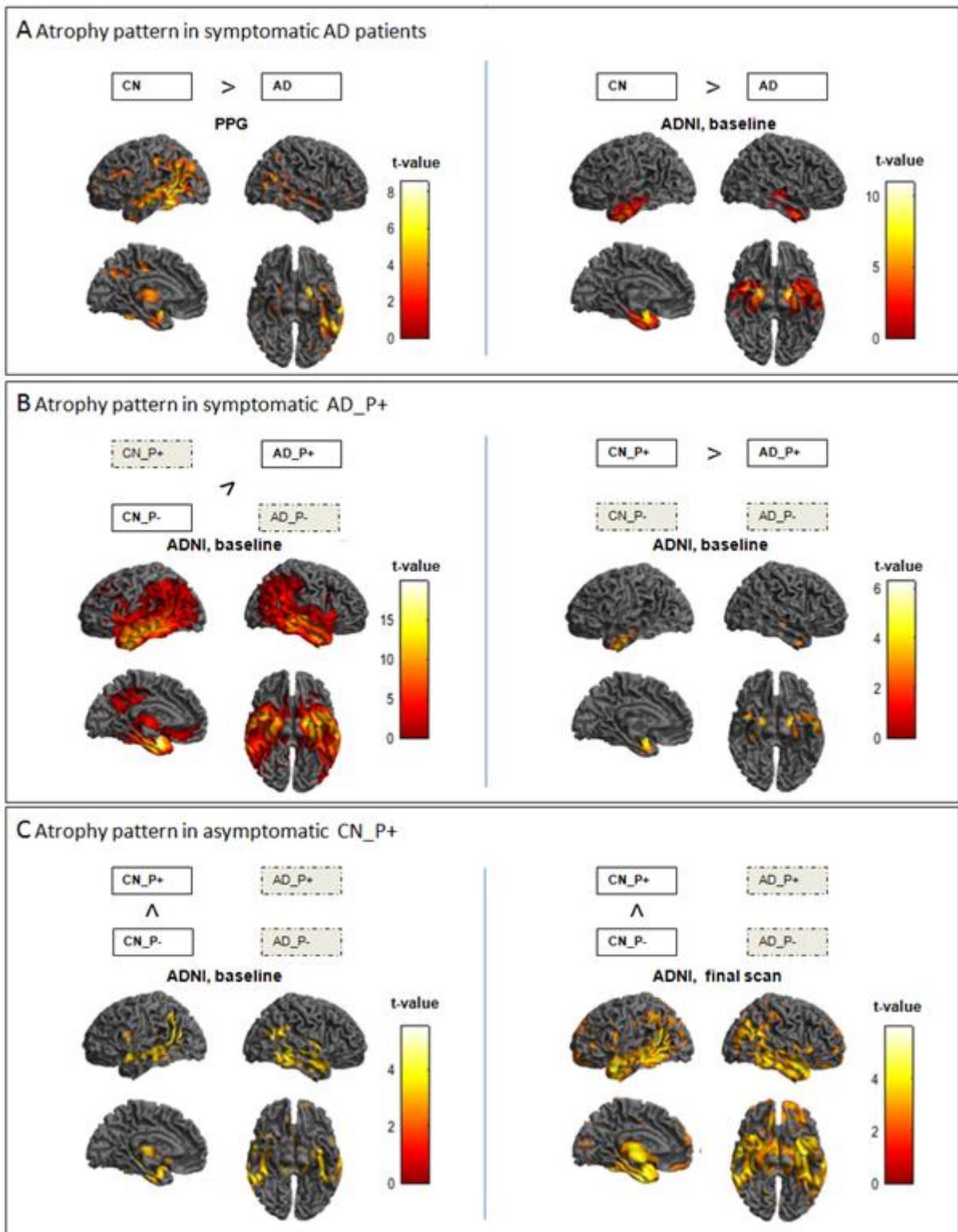


Figure 12 Atrophy pattern across groups from PPG and ADNI cohorts

Panel A shows atrophy pattern of AD compared with CN from pathologically proven group (PPG) and clinically diagnosed ADNI cohort (from left to right).

Panel B shows atrophy pattern of AD_P+ at baseline compared with predicted positive and negative pathology groups of cognitive normal group (CN_P- and CN_P+, from left to right respectively) from ADNI cohort.

Panel C shows atrophy pattern in CN_P+ than in CN_P- from ADNI cohort at baseline (left) and with final scans (right).

Results are shown on three-dimensional renderings of the brain after adjusting of age, gender, TIV (total intracranial volume) and MR strength if available, with $p < 0.001$. Colors show T-score: yellow represents greater volume loss than red. AD=Alzheimer's disease. CN=cognitive normal individuals. AD_P+ = clinical AD predicted pathologically positive. AD_P- = clinical AD predicted pathologically negative. CN_P+ = cognitively normal subjects predicted pathologically positive. CN_P-= cognitively normal predicted pathologically negative.

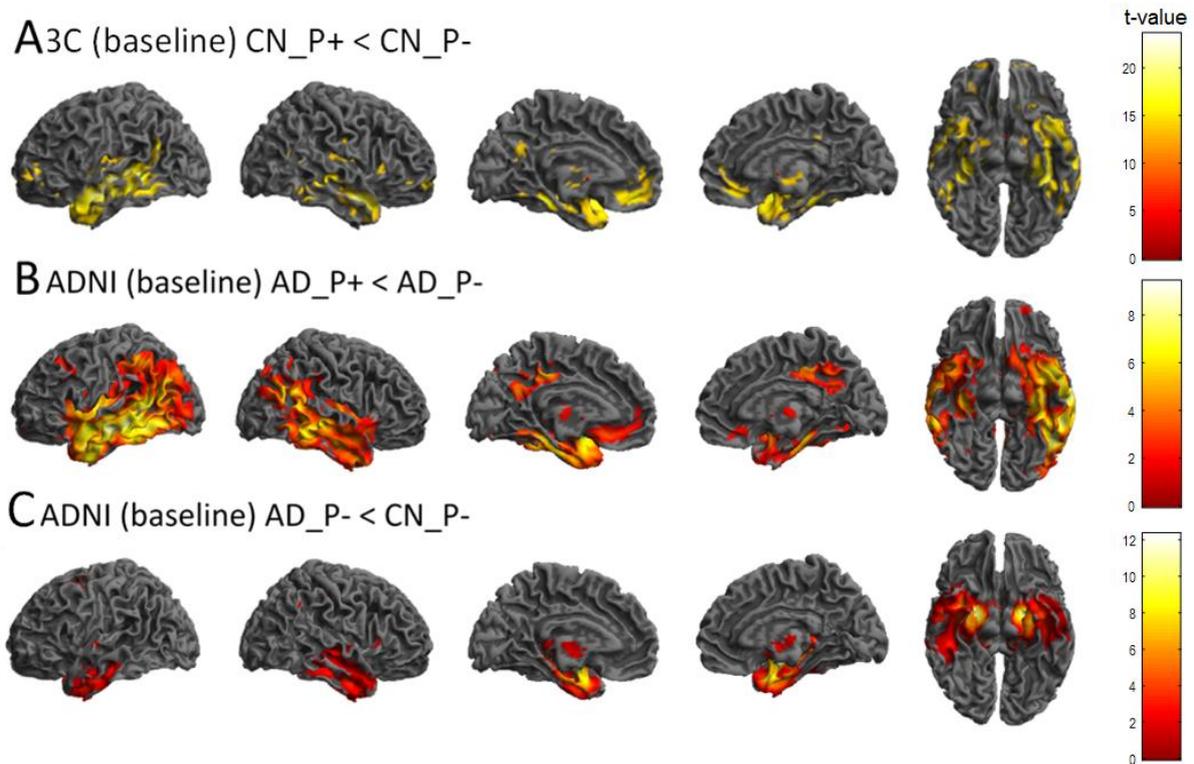


Figure 13 Atrophy pattern across groups from ADNI and 3C cohorts

Panel A shows atrophy pattern of CN_P+ compared with CN_P- from 3C study. ($T > 15$)

Panel B shows relative preservation in ADNI AD_P- compared to ADNI AD_P+. However, this group shows significant atrophy compared with ADNI CN_P- (panel C).

Results are shown on three-dimensional renderings of the brain after adjustment for age, gender, TIV (total intracranial volume) and MR strength, with $p < 0.001$. Colors show T-score: yellow represents

greater volume loss than red. AD_P+ = clinical AD predicted pathologically positive. AD_P- = clinical AD predicted pathologically negative.

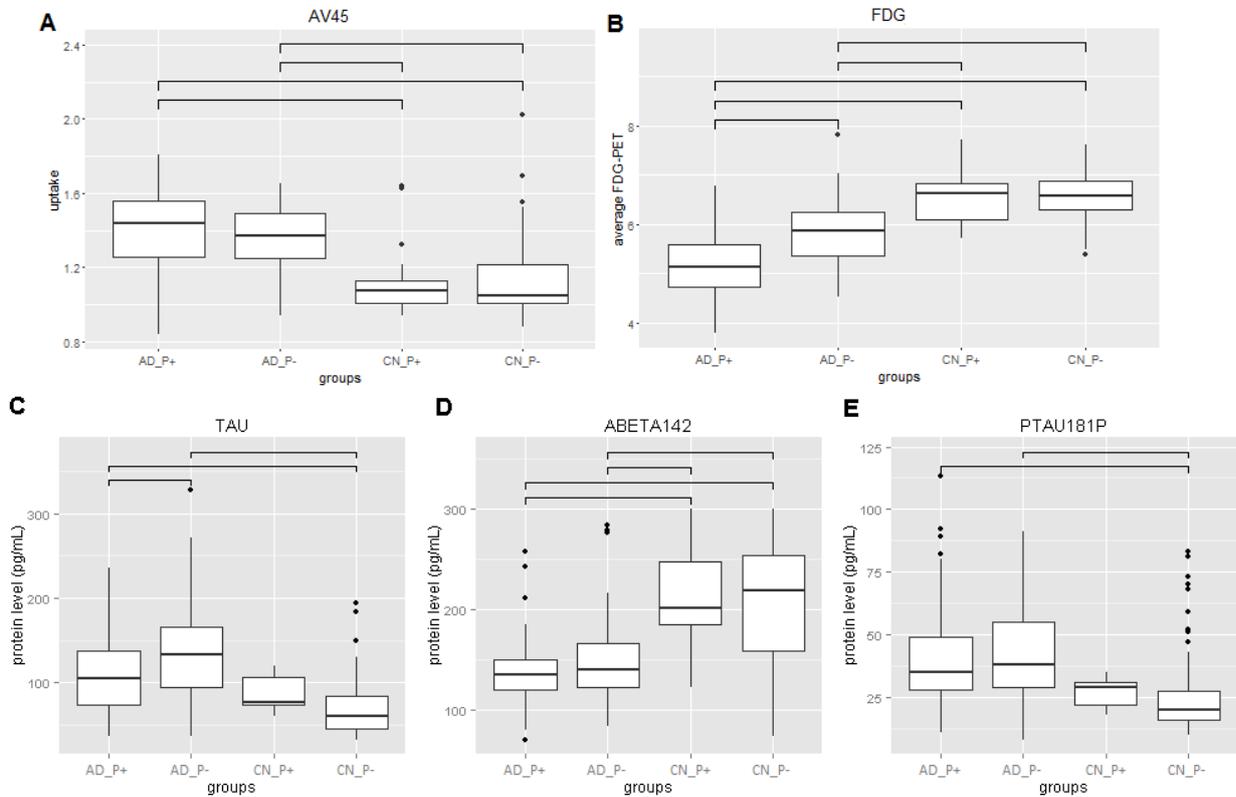


Figure 14 Distributions of biomarkers at baseline between ADNI predicted groups.

Each pair of groups were compared with Wilcoxon rank-sum tests and significant distributions were indicated by a horizontal bar above the relevant groups ($p < 0.05$). For FDG and AV45 examination, the subject numbers are shown in Table 4. For CSF examination, the number of subjects in each group is 57, 41, 5, 108 (ordered as AD_P+ , AD_P- , CN_P+ and CN_P-).

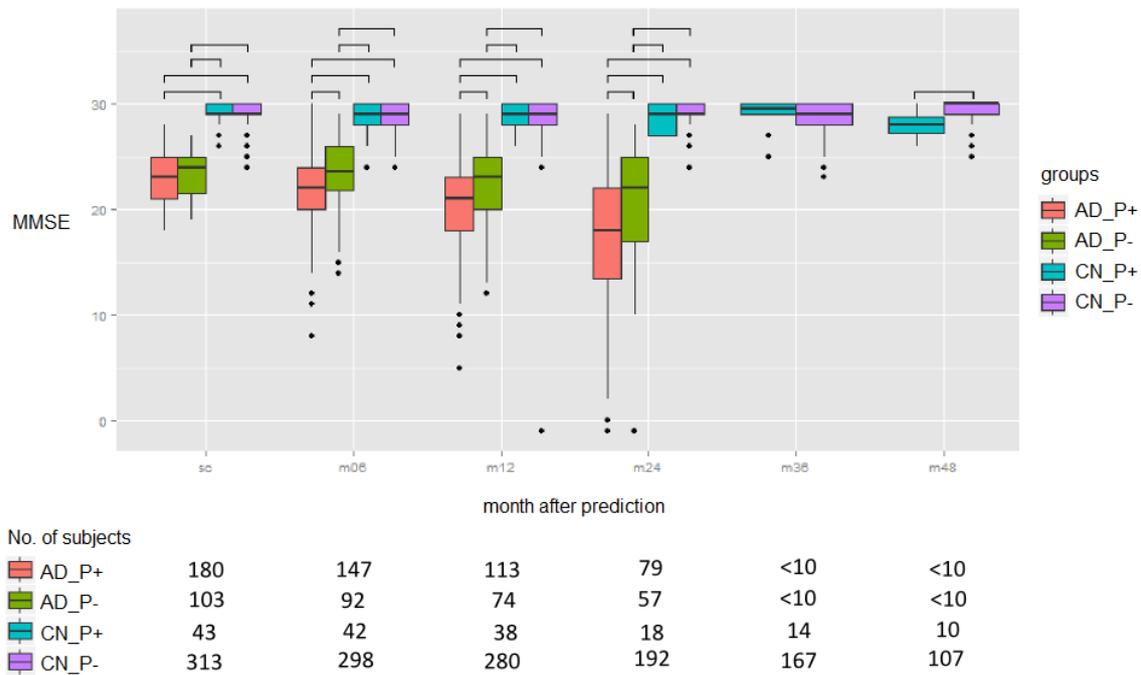


Figure 15 MMSE distributions in follow-up study between ADNI predicted group

Total MMSE scores collected from baseline till 48 months later that are compared between each pair of groups. The statistical test was applied using Wilcoxon rank-sum test and distributions with significant differences were indicated by a horizontal bars above the relevant groups ($p < 0.05$).

4.5 Discussion

This study has demonstrated two important results. Firstly, our automated classifier trained from a clinically and pathologically proven group was able to predict CN_P+ subjects individually from cognitively normal people (CN_P-) in general populations. We showed that this group of individuals are asymptomatic at risk of clinical AD (ASR). Secondly, we detected two groups from clinically diagnosed AD patients (AD_P+ and AD_P-). We observed that they have different atrophy patterns and profiles of biomarkers.

Our results showed that asymptomatic people identified by a pathology based classifier have at least 2.3 times higher risk of clinical conversion with the emergence of symptoms and a fall in MMSE at 4 years after presentation than CN_P- (Figure 15). Therefore, we label this group as ASR. The atrophied regions characterising ASR were replicated in two independent cohorts. These regions include the entorhinal cortex, a region initially affected by Alzheimer's disease according to Braak (stage I). (Braak and Braak, 1995; Khan *et al.*, 2014) At this stage, such individuals are "clinically silent" based on their behavioural symptoms and so not clinical patients, nevertheless significant anatomical change predictive of pathology, is found with MRI-based morphological measurements. At a later

stage, using follow up data from the same individuals, we observed a progression of hippocampal atrophy (Supplementary Figure 1). Therefore, our results match prevalent theories of disease development along limbic stages (Braak stages III and IV), during which many individuals develop impaired cognition and personality changes (Braak and Braak, 1995). This result provides anatomical evidence in vivo that helps understand why ASR (CN_P+) subjects show higher rates of disease conversion. For the AD pathology-related proteins, especially total Tau and pTau from CSF, CN_P- group showed the lowest levels, while CN_P+ group showed notably greater levels. However, the trend did not reach significance due to the small number of subjects (CN_P+ has only 5 subjects).

The proportion of ASR (CN_P+) in both cohorts, 12% of ADNI diagnosed CN and 36% in the 3C study, are comparable to reported clinical misdiagnosis rates that range from 30% to 13%. (Beach *et al.*, 2012) However, the difference between ADNI and 3C populations could be due to differences in diagnostic criteria generating group recruitment biases. The ADNI exclusion criteria were done to exclude participants having no diseases expected to interfere with AD. Compare with ADNI predicted subgroups, 3C CN_P+ is significantly older than CN_P-, indicating that a mild age effect remained even with linear age correction, due to the small sample size in training set. In other words, CN_P+ in 3C might be a mixture of preclinical AD and normal aging. There may also be an effect of known epidemiological differences between enriched research cohorts ADNI and general population 3C studies (Pimouguet *et al.*, 2015).

Among GWAS identified SNPs from previous studies (Carrasquillo *et al.*, 2015; Sleegers *et al.*, 2015), similar to previous results (Carrasquillo *et al.*, 2015; Vivot *et al.*, 2015), none except APOE-ε4 showed significant association with progressive MCI or clinical AD conversion in clinically diagnosed CN. However, we argue that neuroimaging should be used as an endophenotypic marker for genetic association analysis. (Cui *et al.*, 2015) The main advantage of using brain structural changes in this way is that grey matter volume, according to previous studies, accurately predicts Alzheimer's pathology and correlates with Braak stages. (Frisoni *et al.*, 2010; Klöppel *et al.*, 2008)

Compared with AD_P+, AD_P- patients performed better in cognition tests, had higher MMSE score, and showed less grey matter atrophy except in the hippocampus. However, they have significantly higher levels of CSF total tau and glucose metabolism (Figure 14). This intermediate group shows a similar atrophy pattern to a subtype of AD called "limbic-predominant AD" (Whitwell *et al.*, 2012), suggesting a different distribution of neurofibrillary tangles than is usual in typical AD. The atypical AD_P- pattern could be driven by the specificity of our training cohort or be due to a different pathology (for example, primary age-related tauopathy (PART), medial temporal tauopathy or hippocampal sclerosis). A similar group pattern has been found in another study, similar to SNAP (suspected non-Alzheimer disease pathophysiology - Clifford *et al.* (Sperling *et al.*, 2011)), in which it has been argued that the cause may be due to a non-AD degenerative processes.

In summary, the advances of our research rest on two factors. Experimentally, our classifier is trained on data from clinically diagnosed individuals who have that diagnosis confirmed by disease defining pathology. Analytically, we used a multivariate whole-brain voxel-wise approach, which allows the identification of a set of brain regions differentially and specifically affected by AD. This approach differs from a recent study (Stephan *et al.*, 2015), where only 3 global MRI variables were used to predict clinical AD (total white matter lesion volume, total hippocampal volume and total brain volume).

As a preliminary study, our project is limited by its small training set with a risk of having over estimated accuracy. However, the cross-validation accuracy was high (87%) and can be compared with the accuracy from a larger cohort (Klöppel *et al.*, 2008). Our training cohort is special in being closer to the “ground truth” of pathological characterization. It must be admitted there is some way to go before the “ground truth” of cause is identified. As the pathologically verified AD patients were of younger age, the anatomical patterns are less affected by normal aging. Finally, to reduce the chance of over-fitting to our data, we used a linear kernel to generate a simple but sensitive model to identify disease pathology. Other prediction algorithms which go beyond binary answers with a probability of the class assignment of a new subject, like logistic regression, might be more informative and can be tested in the future.

Our findings have significant implications. Our results provide evidence for the appearance of AD-pathophysiology at a preclinical stage before the emergence of clinical symptomatology and speak to the theory of compensatory mechanisms in mature brains, align with fact that pathology precedes by decades before symptoms onset. Our machine-learning based classifier, informed as it was by pathology, could potentially be used in clinical practice to supplement visual rating of hippocampal atrophy and so to improve clinical decision-making and cohort construction for clinical AD trials. We showed evidence of our prediction at baseline and its association with later conversion in CN to clinical Alzheimer’s disease. In reality, the prediction could be done for elderly people regularly to have up-to-date estimation. This can be helpful for providing early treatment. The method may appear complex because it is high dimensional, but applying the classifier to new data is in fact very trivial. Prediction can be obtained by multiplication of the classifier’s weight image by an individual’s MR image. We will make our classifier weight image available so that the method can be applied in clinical practice. It is to be expected that the sensitivity and accuracy of the classifier, which is in principle generic for AD versus non-AD, will increase as more pathologically confirmed AD and CN individuals without AD pathology are added.

We have focused on the ASR group, detected using pathology related topographical markers. Previous studies suggested that CSF markers are more sensitive to early changes than MRI markers. However, the protein levels can vary due to different laboratories and assays (Olsson *et al.*, 2016). Future studies

should compare the sensitivity and specificity of our classifier by using different types of biomarkers, such as Tau imaging (Villemagne *et al.*, 2015). Subgroups like AD_P+ and CN_P- are useful for clinical trials or research studies that need to identify typical AD cases and controls. For AD_P-, further studies with serial clinical observations will be needed with measures of genetic, proteomic and other factors, finally supplemented by pathological examination, to characterize individuals from this atypical AD group.

In conclusion, we provide a prognostic marker for preclinical AD research diagnosis criteria. Among the range of available biomarkers, the choice of a clinically useful ASR subset should be based on discriminative value, cost, the ease and practicality of any procedures involved. In the future, a better model, which reveals the differential causes of subpopulations leading to a common clinical presentation that improve clinical response to a variety of treatments, could be built. To achieve this goal, more data will be needed and could be provided by initiatives such as the medical informatics platform of the Human Brain Project, Dementia Platform UK or aspects of the Innovative Medicines Initiative (IMI2).

5. A data-driven model of estimation a trait to quantify sporadic Alzheimer's disease severity

5.1 Motivation

Alzheimer's disease is a continuous progressive disease. A binary classification is useful in clinical practice but it does not reveal how the brain changes during the disease progression. The benefits of using disease severity quantification are twofold. The estimated severity can be used in clinical practice to supplement visual rating of hippocampal atrophy and so improve clinical decision-making. In addition, the disease severity can be used as an endophenotype, which does not rely on symptoms and has a biological meaning to identify molecular targets for drug development.

To achieve this goal, many solutions have been proposed. A supervised classification method has been tested (Davatzikos *et al.*, 2009; Spulber *et al.*, 2013) over the past seven years. The idea of this method is to determine the brain atrophy features that best distinguish Alzheimer's disease patients from CN in the ADNI study (for example, using support vector machine algorithm detailed in section 2.4). To extract the disease severity, new subjects can be projected onto the weight space. However, this method relies on clinically diagnosed subjects to extract the atrophy features. As introduced in section 3.4 that normal aging has significant confounding effects and affects regions known to shrink in AD patients, therefore the features are not necessarily associated with disease. To reduce age-related effects in modeling neurodegenerative brain disease, particular attention needs to be paid to the sample selection in terms of age between controls and cases.

A data driven method has been recently proposed (Young *et al.*, 2014a) which integrates an event-based model and Markov chain Monte Carlo algorithm to estimate a sequence of events that maximizes the data likelihood. More details are explained in section 2.4.6. The event is defined as a switch of one biomarker from normal to abnormal. The input of the algorithm includes a list of biomarkers and risk factors, such as A β 42 and Tau proteins from CSF, grey matter volume of hippocampus and entorhinal cortex and clinical measurements such as the MMSE score. The stage will be assigned which maximizes the probability of the data given the maximum likelihood event sequence.

There are two limitations to this method. Firstly, the algorithm is designed in a way that the input requires a list of risk factors. These risk factors can be identified from previous studies. Therefore the selection of the risk factors influences the evaluation of the staging estimation. The second limitation is the calculation of each event: a risk factor switch from normal to abnormal state. The algorithm integrates a Gaussian mixture model on each risk factor to stratify the subjects into two states. However, this step ignores the fact that Alzheimer's disease and related risk factors progress

continuously rather than switching from one state to another. Therefore, an extension of such a model by using continuous variables is a step closer to reality in the quantification of disease severity.

5.2 Introduction

Several methods are used to estimate Alzheimer's disease (AD) severity *in vivo*. The most widely used indicators of disease severity include The Mini-Mental State Examination (MMSE) and the Global Score of the Clinical Dementia Rating Scale (CDR) (Morris, 1993). However, clinical tests do not reveal the biological change in the brain (Dubois *et al.*, 2014).

Alzheimer's disease severity and Braak stages are defined with respect to the distribution of neurofibrillary tangles in the brain (Braak and Braak, 1995). This can only be confirmed through a post-mortem examination. Dubois has proposed in IWG-2 in 2014 by to use topographical or downstream markers to indicate clinical severity (Dubois *et al.*, 2014). Topographical or downstream markers measure Alzheimer's pathology-related metabolic changes or neuronal loss. Therefore they identify downstream brain changes indicative of the regional distribution of Alzheimer's pathology, for example medial temporal lobe atrophy on MRI (Dubois *et al.*, 2014). The benefits of using an estimated continuous severity are two-fold, to monitor dementia onset and to quantify disease stages across Alzheimer's disease spectrum.

The grey matter loss patterns detected by MR scans correlate with neurofibrillary tangles (Braak and Braak, 1995; Whitwell *et al.*, 2008). Medial temporal atrophy was considered to be the best MRI marker for further clinical progression at a prodromal (preclinical) stage, and hippocampal atrophy has been proposed to be the most robust (Dubois *et al.*, 2014). But the MRI biomarker of hippocampal atrophy has been criticized as its volumetric change can be caused in several non-AD conditions.

An alternative idea of summarizing biomarkers into one trait to quantify disease severity has been investigated in several studies. Two types of algorithms have been studied to estimate disease severity. These approaches are based on classification algorithms using atrophy pattern (Davatzikos *et al.*, 2009) or unsupervised algorithms, such as the event-based model (Young *et al.*, 2014b). However, these methods rely on clinically diagnosis to detect disease-related regions or on a set of selected biomarkers.

In this study, we propose to extract a latent variable based on MRI-measured atrophy patterns to quantify the disease severity for each subject by applying item response theory (IRT). IRT has been well used in psychological tests to identify latent traits, such as personality and mood traits (Ferrando, 2002). To improve on the limitations of previous studies, we propose to apply an IRT model for continuous data, continuous response model, based on factor analysis using grey matter volumes to identify a disease-related latent trait or a disease-related latent factor and its relation with grey matter

regions. The main advantage of this method is that it does not completely rely on the supervision of clinical diagnosis to identify disease-related regions.

The aim of this study is to test whether the estimated latent trait can be used to quantify the disease stage and predict the risk of disease onset. Its relevance to AD has been demonstrated through several lines of evidences. We showed that the latent trait associates with clinical diagnosis, the performance of clinical tests at baseline and longitudinal clinical conversion. It strongly correlates with CSF / cortical $A\beta$, CSF Tau and phosphorylated Tau levels, which have been found to associate with cortical amyloid deposition, intensity of neurodegeneration and neurofibrillary pathological changes (Scheltens *et al.*, 2016). In addition, we also compared the latent trait to the $A\beta$ /Tau ratio from CSF, which has been tested in several studies for its predictive power in longitudinal AD development (Jack *et al.*, 2013). The estimated latent trait's associations with glucose metabolic reductions and APOE- ϵ 4 were also reported.

5.3 Method

5.3.1 Data

Two independent cohorts have been used in this study i) Alzheimer's Disease Neuroimaging Initiative (ADNI) (Petersen *et al.*, 2010) data , ii) the French population-based Three-City (3C) Study, Dijon cohort. (3C Study Group, 2003; Alperovitch *et al.*, 2002).

1) ADNI

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 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). For up-to-date information, see www.adni-info.org .

The data used from ADNI database, 28th of Nov, 2014, comprises T1-weighted MR scans from clinically diagnosed individuals acquired at baseline, including 322 cognitive normals (CN), 62 subjective memory complaints (SMC), 693 mild cognitive impairments (MCI, 217 early MCI (EMCI) and 476 late MCI (LMCI)) and 252 AD. Images with artifacts due to motion were removed in visual quality control. Baseline clinical tests information, including MMSE (Mini-Mental State Examination, 1267 subjects), CDR-SB (Clinical Dementia Rating scale Sum of Boxes, 1267 subjects) and ADAS (Alzheimer's Disease Assessment Scale, 1257 subjects) was downloaded. Longitudinal diagnostic

information of CN and MCI subjects were collected. CSF markers of Alzheimer's pathology, such as Tau, A β 42 and pTau concentrations measured at baseline, were downloaded. In total, 369 subjects had Tau protein information; 374 had A β 42 and 375 had pTau. The log transformation with natural base was applied on the clinical tests and the proteins to improve normality. Cortical A β 42 burden was measured by florbetapir F 18 (also known as [18F] AV45) ('ADNI procedures manual', n.d.). The level of A β 42 burden is the average AV45 standard uptake value ratio of frontal, anterior cingulate, precuneus, and parietal cortex relative to the cerebellum at baseline. The severity of hypometabolism from [18F]2-fluoro-2-deoxyglucose (FDG) PET was the average FDG-PET of angular, temporal, and posterior cingulate at baseline ('ADNI procedures manual', n.d.). APOE- ϵ 4 genotype, defined by rs429358 and rs7412, were measured for 1322 subjects.

2) 3C

This study uses the first recorded MR images from the clinically diagnosed 1447 aged CN of the French 3C study at baseline, Dijon cohort (Table 1). All subjects' longitudinal diagnostic information were downloaded. CN subjects who developed symptoms related to other clinical types of dementia in the follow up study were excluded. Study details are described in Stephan et al. (Stephan *et al.*, 2015).

5.3.2 Preprocessing

Grey matter volumes were measured in all study individuals using SPM12, an open source software package written in Matlab (<http://www.fil.ion.ucl.ac.uk/spm/software/spm12/>).

The T1-weighted images were automatically segmented into 114 anatomical structures using the Neuromorphometrics atlas (the full list of the structures can be found in Supplementary Material). The volumetric data is M x N, where 'M' rows represent the number of subjects, and 'N' the number of anatomical structures. In short, the methodology consists of two main steps. Firstly, each individual T1-weighted image is normalized to MNI (Montreal Neurological Institute) space using non-linear image registration Shoot toolbox (SPM12). Additionally, in this step the individual images are segmented in three different brain tissues (cerebral spinal fluid, gray matter and white matter). Secondly, each individual gray matter voxel is labeled based on Neuromorphometrics atlas (constructed by manual segmentation for a group of subjects) and the transformation matrix obtained in the previous step. 'Neuromorphometrics' maximum probability tissue labels were derived from the 'MICCAI 2012 Grand Challenge and Workshop on Multi-Atlas Labeling' (https://masi.vuse.vanderbilt.edu/workshop2012/index.php/Challenge_Details). These data were released under the Creative Commons Attribution-Non-Commercial (CC BY-NC) with no end date. The MRI scans originate from the OASIS project (<http://www.oasis-brains.org/>) and the labelled data

was provided by Neuromorphometrics, Inc. (<http://Neuromorphometrics.com/>) under academic subscription.

5.3.3 Statistical analysis

1) Feature scaling

To avoid inter-subject and inter-cohort variance, each regional volume (V_{Region}) was scaled by the total grey matter volume ($V_{TotalGreyMatter}$) (equation (5.1)). This step corresponds to the first step in item response theory, which determines the proportion of items that each respondent answered correctly (Baker, 2001). The ratios between V_{Region} and $V_{TotalGreyMatter}$ were used as limited responses.

$$\mathbf{ratio}_{regional} = \frac{V_{Region}}{V_{TotalGreyMatter}} \quad (5.1)$$

To apply a continuous response model, the ratios were firstly transformed with a logit function:

$$V_{regional} = \mathbf{logit}(\mathbf{ratio}_{regional}) = \log\left(\frac{\mathbf{ratio}_{regional}}{1 - \mathbf{ratio}_{regional}}\right) \quad (5.2)$$

into continuous-unlimited responses (Lesaffre *et al.*, 2007). Secondly, a linear regression model was constructed to remove the effects of a number of covariates. The covariates included age, gender, total intracranial volume (TIV) and magnetic resonance strength field in ADNI cohort. In 3C cohort, magnetic resonance strength field was ignored as all subjects were scanned with 1.5 tesla scanners. The adjusted logit transformed ratios (V) were then used in linear factor analysis to estimate the latent trait.

2) Linear factor analysis (FA) and item characteristic function

In psychometric studies, item response theory is widely used to estimate a latent trait or a latent factor, which cannot be measured directly by psychometric tests. When the responses are in continuous formats, factor analysis, a congeneric model can be applied (Ferrando, 2002; Ghahramani and Hinton, 1997). In this model, a latent trait and the responses can be expressed as:

$$\mathbf{V} = \mathbf{C} + \mathbf{FH} + \mathbf{E}, \quad (5.3)$$

where V is a $D \times N$ matrix of the observed responses, in our case the adjusted ratios of regional volumes. D denotes the number of brain regions and N the number of subjects. C has the same dimension as V with each row the mean of one brain region across N subjects. F is termed as the factor loading matrix, where each row is a response variable and each column is a factor with f factors

in total. F represents the relation between the latent traits and the observed responses (shown in (8.1)). Matrix H , with dimension $f \times N$, represents f latent factors for N subjects and each factor is assumed to be $N(0, I)$. Each row of the residual term, E , is assumed to be normally distributed with a diagonal covariance matrix Ψ . Thus, the observed data V follows the distribution of $N(c, FF^T + \Psi)$, where $c \in R^D$ is one column of C . The goal of factor analysis is to estimate H, F and Ψ that best model V . Expectation maximization algorithm was used to estimate F and Ψ . This algorithm was finding parameters, F and Ψ , that maximize the (log) likelihood of data V through an iterative procedure. The initial guess of the parameters was estimated using principal component analysis. Steps are shown in Supplementary Material. For a given F , H can be estimated through the least squares fitting technique, which is most commonly applied in linear regression.

Item characteristic function (ICF) is defined as the regression of the item response score, V in this project. In continuous response model, ICF is the conditional distribution of V for fixed H . In this project, ICF represents the change of the ratios of regional volume across the estimated latent trait which we expect it to quantify disease severity. Details are shown in Supplementary Material.

3) Linear FA estimation

Using linear FA on the adjusted transformed ratios, we obtained the first three underlying factors and their loadings through the EM algorithm. Due to the fact that ADNI data is composed of cognitive controls, MCI and AD subjects, we assumed that the biggest variance of the data relates to Alzheimer's disease and the first factor is a disease-related latent trait. After applying linear FA on ADNI cohort, the estimated factor loadings were used for estimating the factors for the 3C subjects by with the least-square approach.

4) Association with baseline clinical diagnosis, follow-up clinical conversion and biomarkers

All statistical tests were performed in R. To determine whether the estimated latent factor is useful in quantifying disease severity, its associations were compared with clinical and biological measurements. The Wilcoxon rank-sum test was used to compare the association with baseline clinical diagnosis. Cox proportional hazards regression models were used to identify the association between AD risk factors (such as the estimated latent trait and age) and event-free survival time in follow-up studies of ADNI and 3C cohorts. Event-free survival is defined as the time of CN from baseline until the first clinical diagnosis of MCI, when available, or AD. For MCI group diagnosed at baseline, the event-free survival is the time until the first clinical diagnosis of AD. As the relation between SMC and Alzheimer's disease has not been fully understood (Sajjad *et al.*, 2015), SMC group was not used for comparison with clinical conversion but was used to compare with other markers. The latent trait

was tested for its association by calculating the Spearman's correlation coefficients with: CSF biomarkers, including three CSF proteins (A β 42, Tau and pTau) and A β 42/Tau ratio; cortical A β 42 burden measured from AV45 PET; glucose metabolic reductions estimated from FDG PET; three baseline clinical tests, MMSE, CDRSB and ADAS scores. Cox proportional hazards regression models were constructed with the R package *survival* (Therneau and Lumley, 2016).

5) Factor loadings and item characteristic function

To identify whether the grey matter regions affected by Alzheimer's disease have higher correlation with the estimated latent trait than other regions, all grey matter regions were ranked according to their factor loadings. A linear model, using each of the grey matter voxels as a dependent variable, was built in order to visualize which regions associate with the estimated latent trait. This model was adjusted for the covariates: age, gender, TIV and MR strength. The expected regional ratios given latent traits were estimated using an item characteristic function (details are shown in Supplementary Material and (Ferrando, 2002)). The item characteristic curves for the top 6 highly associated grey matter regions are given.

5.4 Results

5.4.1 The estimated latent trait and its association with clinical diagnosis

Figure 16 A shows the latent trait of all ADNI subjects ranked based on their latent trait. The probability density distribution and cumulative density distribution of the estimated latent trait of all subjects from ADNI cohort is given in Figure 16 B. The probability density distribution shows the probability that the estimated latent trait has a given value. The cumulative density distribution shows the probability that the estimated latent trait has a value smaller than a given value. The density distribution for each clinical group in box plot is given in Figure 16 C. A significant association was observed between the estimated latent trait and the baseline clinical diagnosis (CN vs EMCI, $p=4.855e-11$; EMCI vs LMCI, $p=0.01252$; LMCI vs AD, $p<2.2e-16$, Wilcoxon rank-sum test). The Spearman's correlation coefficients between the estimated latent trait and the clinical tests are 0.43 ($p=6.73e-59$, through Fisher's Z-transformation), 0.4 ($p=1.4e-49$, through Fisher's Z-transformation) and 0.47 ($p=1.68e-69$, through Fisher's Z-transformation) for CDR-SB, MMSE and ADAS scores (Figure 16 D, E, F).

In addition, a significant association was found between the estimated latent trait and the event-free survival time in follow-up studies, after adjusting for the potential confounders, including APOE- ϵ 4, age and education, in a Cox model (Table 8). For ADNI CN group, two types of event-free survival time were tested. The first type was defined as the time of CN from baseline until the first clinical diagnosis of MCI, when available, or AD (model 1, HR=1.61, 95% CI 1.1 to 2.3, $p=0.009$, Wald test).

The second type was defined as the time of CN from baseline until the first clinical diagnosis of MCI (CN subjects converted to AD were excluded) (model 2, HR=1.52, 95% CI 1.04 to 2.22, $p=0.03$, Wald test). For 3C cohort, the event was defined as clinical AD (model 3, HR=1.59, 95% CI 1.1 to 2.2, $p=0.0069$, Wald test). For ADNI MCI group, including EMCI and LMCI, the event was defined as clinical AD (model 4, HR=1.75, 95% CI 1.5 to 2, $p=2.32e-14$, Wald test). Pathology related variables, CSF proteins level, were added in an independent Cox model for testing whether the latent trait can predict AD beyond CSF proteins. The latent trait showed significant association with MCI-free survival for CN group and clinical AD-free survival for MCI group (results are shown in Supplementary Table 1).

5.4.2 The estimated latent trait and its association with protein, metabolism levels and APOE- ϵ 4 genotype

The estimated latent trait significantly associated with three CSF proteins, pathology related biomarkers in Alzheimer's disease. Spearman's correlation coefficients were: 0.34 ($p=1.47e-11$, through Fisher's Z-transformation), -0.33 ($p=6.28e-11$, through Fisher's Z-transformation) and 0.32 ($p=2.1e-10$, through Fisher's Z-transformation), for Tau, A β 42 and pTau separately. Apart from the three CSF proteins, the correlation between the estimated latent trait and a ratio between A β 42 and Tau was tested, since in previous studies this ratio has been found to be predictive to longitudinal AD development. The result shows a higher correlation to the ratio than to each single CSF protein (-0.38, $p=3.35e-14$, through Fisher's Z-transformation) (Figure 17 A, B, C, D). Cortical A β 42 burden levels measured by AV45 PET significantly associated with the estimated latent trait with a correlation coefficient equals to 0.21 ($p=2.6e-06$, through Fisher's Z-transformation) (Figure 17 E). A significant correlation was also observed between the latent trait and the blood glucose level ($\rho=-0.35$, $p=7.94e-27$, through Fisher's Z-transformation) (Figure 17 F). For subjects carrying different copies of APOE- ϵ 4, their latent traits showed significantly different distributions ($p=1.825e-08$ and 0.0002 for 0 vs 1 and 1 vs 2 respectively, Wilcoxon rank-sum test) (Figure 18).

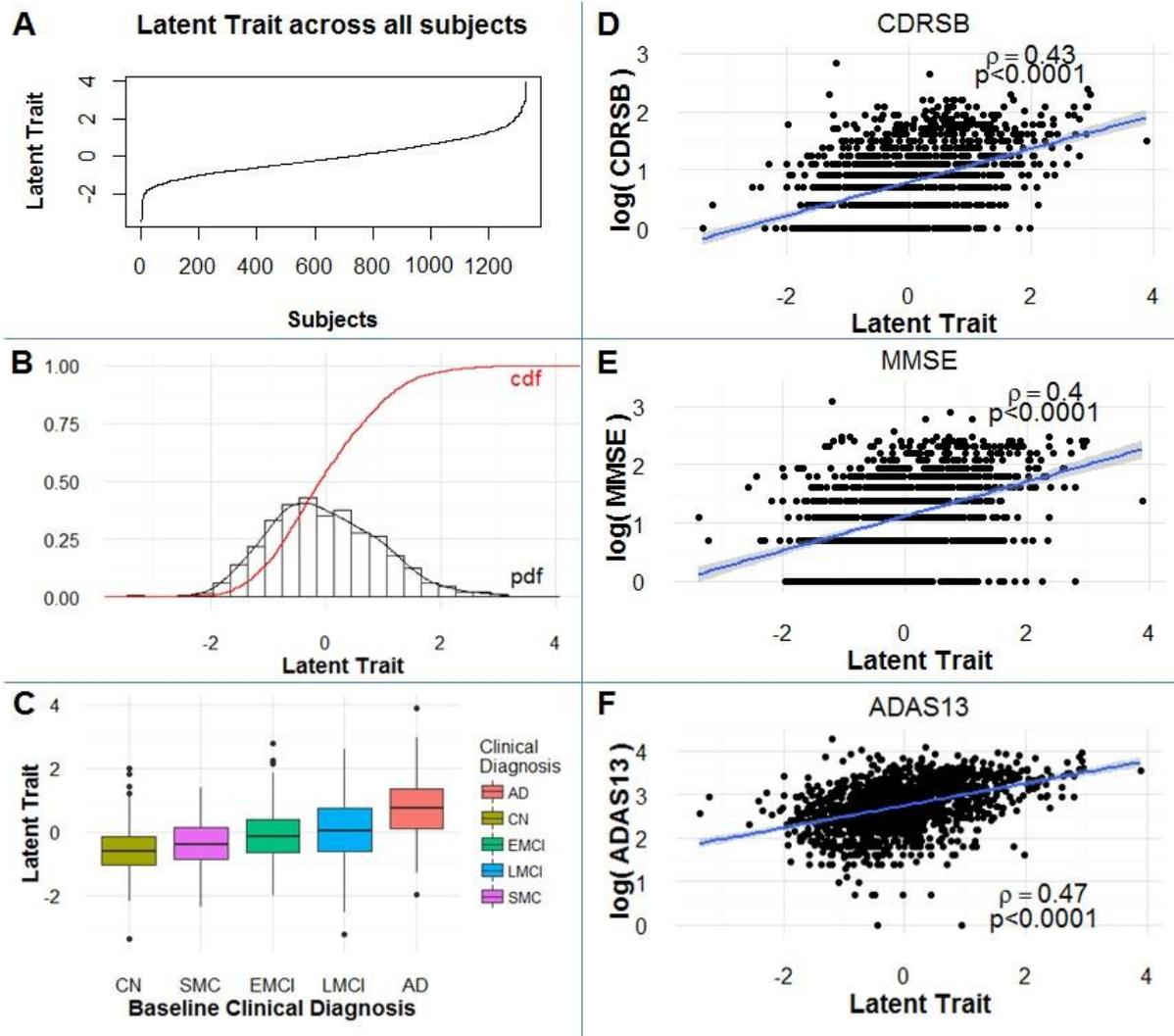


Figure 16 Association between the latent trait and clinical measurements in ADNI cohort

Panel A shows the ranked latent trait values of ADNI subjects

Panel B shows probability density and cumulative distributions of the estimated latent trait from ADNI

Panel C shows the distributions of the latent trait of all clinical diagnosed groups

Panel D, E and F are scatter plots for the log transformed clinical test scores with the latent trait, using subject samples from ADNI cohort. Spearman correlation is calculated. The fitted regression line from simple linear model is plotted in blue, and the standard error is shown in grey. P-value indicates whether the correlation coefficient is significant.

	<i>ADNI cohort</i>					<i>3C cohort</i>
	Cognitive Normal	Subjective Memory Complaints	Early Mild Cognitive Impairment	Late Mild Cognitive Impairment	Alzheimer's Disease	Cognitive Normal
	N=322	N=62	N=217	N=476	N=252	N=1447
Age- yr^a	75.04±5.66	71.67±5.57	70.44±7.26	73.89±7.66	75.1±7.72	72.62±4.09
Male sex – no. (%)	159(49%)	26(42%)	116(53%)	285(60%)	130(52%)	561(38.8%)
Education – yr^a	16.27±2.66	16.73±2.61	15.99±2.68	15.89±2.91	15.14±2.94	3.95±1.57
APOE-ε4 carrier (%)	90(28%)	20(33%)	93(44%)	248(53%)	169(68%)	323(22%)
APOE-ε4 1 copy (%)	79(25%)	20(33%)	83(39%)	183(39%)	115(46%)	
APOE-ε4 2 copy (%)	11(3%)	0	10(5%)	65(14%)	54(22%)	

Table 7 ADNI and 3C studies baseline demographic information

^a Plus-minus values are means ± SD.

variables	Model 1 ^b ADNI CN to MCI or AD		Model 2 ^c ADNI CN to MCI		Model 3 3C CN to AD		Model 4 ADNI MCI to AD	
	Hazard Ratio (95% CI)	P Value	Hazard Ratio (95% CI)	P Value	Hazard Ratio (95% CI)	P Value	Hazard Ratio (95% CI)	P Value
Latent Trait	1.61 (1.13-2.31)	0.009	1.52 (1.04-2.22)	0.03	1.59 (1.14-2.22)	0.007	1.75 (1.51-2.01)	p<0.0001
Female	0.74 (0.43-1.26)	0.264	0.76 (0.43-1.34)	0.341	0.77 (0.47-1.26)	0.298	1.02 (0.78-1.33)	0.886
Age	1.02 (0.97-1.07)	0.396	1.01 (0.96-1.06)	0.758	1.2 (1.13-1.28)	p<0.0001	1.02 (1-1.04)	0.044
APOE-ε4 dose								
2	6.43 (2.23-18.51)	0.001	1.96 (1.09-3.53)	0.024			2.19 (1.48-3.23)	p<0.0001
1	1.83 (1.05-3.17)	0.032	7.24 (2.49-21.09)	0.0003			1.92 (1.45-2.54)	p<0.0001
1 or 2^a					2.57 (1.6-4.12)	p<0.0001		
Education	0.91 (0.82-1.01)	0.081	0.93 (0.83-1.04)	0.187	0.84 (0.72-0.98)	0.024	0.99 (0.95-1.03)	0.588

Table 8 Effect of risk factors on conversion from cognitive normal to MCI/AD and from MCI to AD in ADNI and 3C cohorts

^a 3C cohort has no information about the number of APOE-ε4 allele

^bEvent-free survival is defined as the time of CN from baseline until the first clinical diagnosis of MCI, when available, or AD.

^cEvent-free survival is defined as the time of CN from baseline until the first clinical diagnosis of MCI.

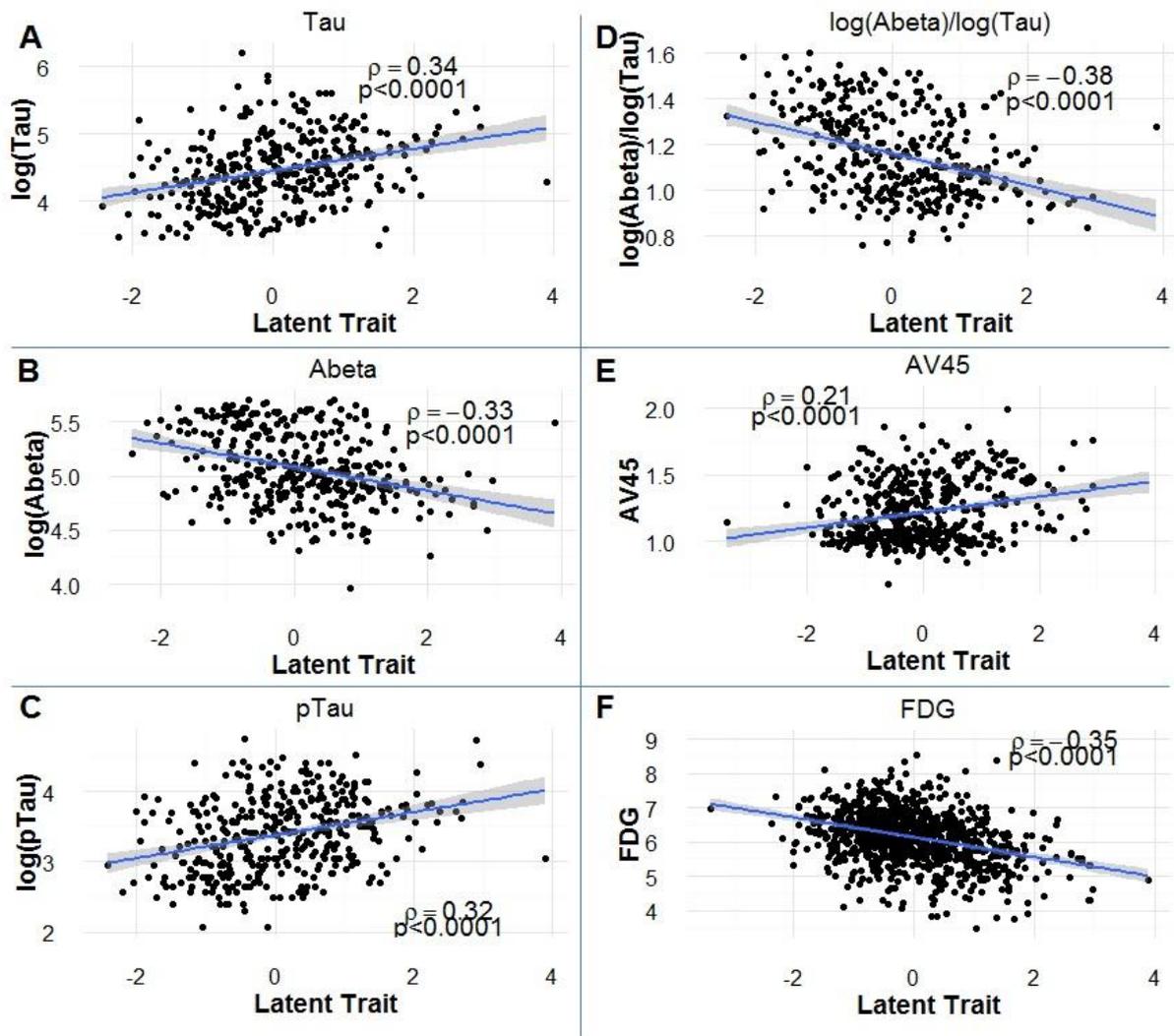


Figure 17 Association between the latent trait and biomarkers in ADNI cohort

Panel A, B and C scatter plots for the log transformed CSF protein levels with the latent trait, using subject samples in ADNI cohort. Spearman correlation is calculated.

Panel D is a scatter plot of the ratio between log transformed CSF protein levels, $A\beta_{42}$ and Tau, as a function of the latent trait (Spearman correlation), for subject samples in ADNI cohort.

Panel E shows the correlation between cortical $A\beta_{42}$ burden and the latent trait, for subjects from ADNI

Panel F shows the correlation between glucose uptake level and the latent trait, for subjects from ADNI

The fitted regression line from simple linear model is plotted in blue, and the standard error is shown in grey. P-value indicates whether the correlation coefficient is significant.

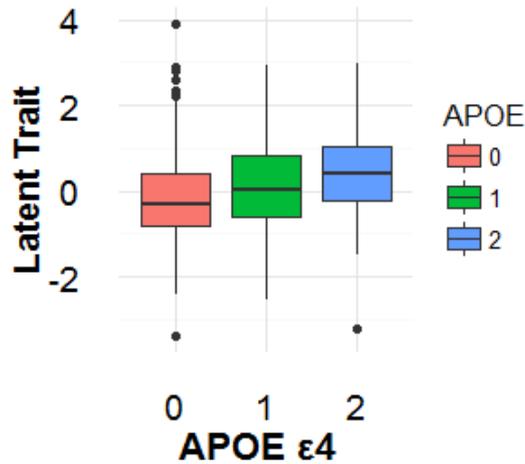


Figure 18 Box plot of the estimated latent trait for APOE-ε4 groups

This figure shows the distributions of the estimated latent trait for subjects with 0, 1, and 2 APOE-ε4 alleles. Groups 0, 1 and 2 in APOE-ε4 have significantly different distributions ($p=1.825e-08$ and 0.0002 for 0 vs 1 and 1vs 2 respectively, Wilcoxon rank-sum test).

5.4.3 Item response theory to estimate regional change

The grey matter regions were ranked based on their correlation with the latent trait, indicated as F (equation (8.1), Supplementary Figure 3). The highly associated regions are from both temporal lobes especially (amygdalae, entorhinal cortices, hippocampi and parahippocampal gyri) (Supplementary Figure 4, one voxel example is shown in Supplementary Figure 5). These regions showed significant negative association with the latent trait, after adjustment for disease related confounders such as age, gender, TIV and MR strength (Figure 19). Regions affected by Alzheimer's disease monotonically decreased across the latent trait in ratio of volume to the total grey matter volume. However, the regions which are not expected to be affected by Alzheimer's disease, such as frontal pole areas, inferior, medial and superior frontal gyri showed a monotonically increase across the latent trait in ratio of volume to the total grey matter volume (Supplementary Figure 6).

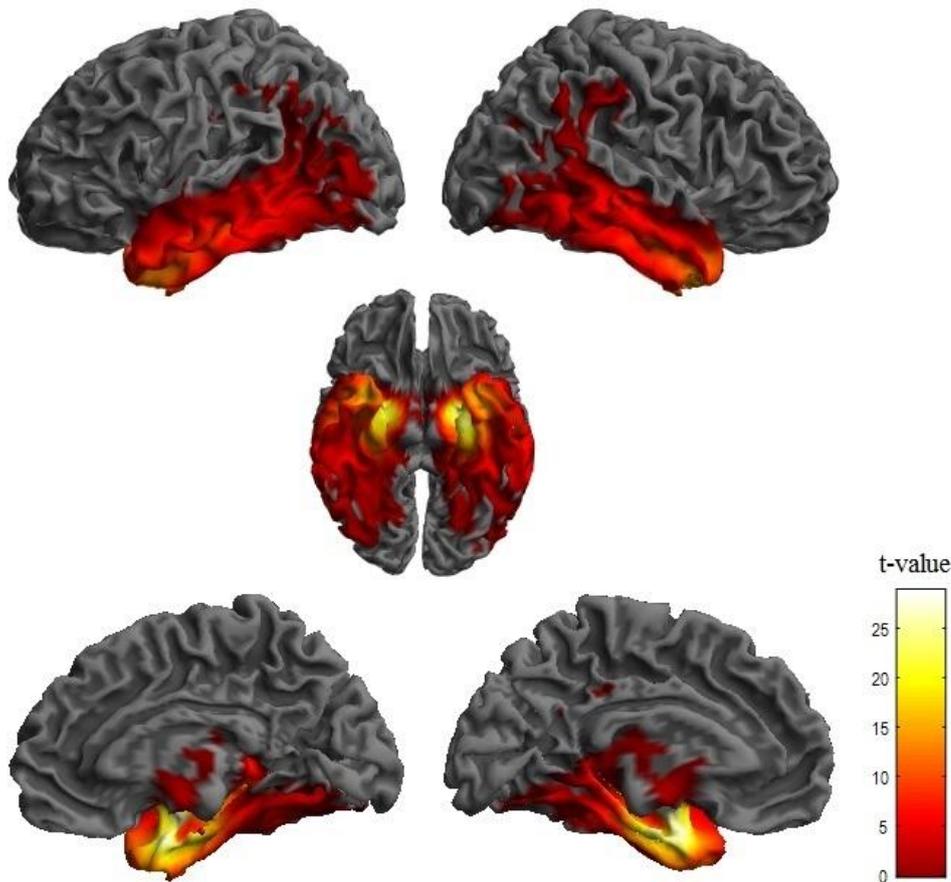


Figure 19 Regions strongly associated with the estimated latent trait

Results are shown on three-dimensional renderings of the brain, the association was tested after adjustment of age, gender, TIV (total intracranial volume) and MR strength, with $p < 0.05$ (FWER). Colors show T-score: yellow represents greater volume loss than red.

5.5 Discussion

This is the first study adapts the continuous response model to estimate a latent trait based on neuronal loss. The latent trait significantly associates baseline clinical diagnosis and tests, follow-up clinical conversion in two independent cohorts, protein levels from CSF and cortical regions, metabolism and APOE- $\epsilon 4$ genotype. Regions affected by Alzheimer's disease are highly associated with the latent trait. Overall, the results provide evidence in support of using the latent trait in quantifying disease severity.

The estimated disease severity showed a significant association with baseline clinical diagnosis. However, the different clinical groups show largely overlapping latent trait values (Figure 17 C). Our interpretation is that the clinical labels were assigned mainly based on symptoms, which do not necessarily reveal the biological change in the brain. Many studies have shown that the disease

precedes clinical symptoms and there is a long asymptomatic period (Braak and Braak, 1995; Dubois *et al.*, 2016; Jack *et al.*, 2014). Several variables have been identified to have influence on brain structure or clinical symptoms. Disease-related cofounders, such as normal aging, have shown a significant effect on regions affected by AD (e.g. hippocampus, entorhinal cortex) (Huijbers *et al.*, 2014). In this study, we did not observe any linear or non-linear correlation between age and the estimated disease severity in both cohorts. In the ADNI cohort no correlation between age and the disease related trait has been identified. Education has been found to reduce the influence of the pathology on the clinical expression of symptoms (Brayne *et al.*, 2010). The exact roles of these variables on Alzheimer's disease are still unclear. These variables can bring variations into the data even though the covariates were adjusted before the linear FA.

We have shown that the estimated disease severity associates with disease conversion and pathology-related markers. The proportional hazards of clinical conversion among cognitive normal groups are consistent from two independent cohorts, ADNI and 3C, with HR values around 1.5. From a biological mechanism point of view, the estimated disease severity has strong correlations with pathology-related protein levels from CSF and cortical regions. These proteins are the hallmarks of Alzheimer's disease in autopsy examinations and have been shown to correlate with AD amyloid / neurofibrillary pathology in autopsy examination (Buerger *et al.*, 2006; Doraiswamy *et al.*, 2014; Strozyk *et al.*, 2003). Due to the high accessing complexity and cost reasons, our method could be used as a surrogate of the proteomic measurements in clinical practice. In addition, the regions with higher loadings are those affected by Alzheimer's disease, according to several previous studies (Braak and Braak, 1995; Huijbers *et al.*, 2014; Sperling *et al.*, 2011). These results support the hypothesis that the estimated disease severity can be used to quantify disease severity.

Summarizing, this study advances the research on two fronts. The method we propose to quantify the disease severity is based on brain topographical markers, which indicate regional distribution of Alzheimer's pathology. Therefore it has biological meaning and is not biased by symptoms. Analytically, a multivariate whole-brain approach was used, which allows the identification of a set of brain regions without relying on the disease status of the participants.

As a preliminary study, our project is limited by its dataset, the ADNI study, which still relies on clinical diagnosis. Although our method requires the input data set are related to AD, nevertheless it is unlike the supervised algorithms which completely rely on the clinical diagnosis to identify the disease-related regions. To improve the specificity of the data, subjects who had significant neurological disease other than Alzheimer's disease were excluded during the acquisition procedure. This study has only focused on Alzheimer's disease. Other types of neurodegenerative diseases with similar atrophy patterns during a specific period, such as tauopathy, are undistinguishable. This drawback can be solved in the future when MRI data on multiple neurodegenerative diseases become

available. Then the severity and relationship between the atrophy patterns via multiple latent factors, each of which indicating the severity of a particular disease, can be quantified.

Our findings have significant implications. Our method evaluates the severity of each subject in the AD spectrum to supplement the visual rating of hippocampal atrophy and assign each participant into a clinical group. For those who are at a prodromal (preclinical) stage, the severity can also provide a relative risk of clinical conversion to symptomatic stages. In clinical practice, the severity can be used to improve clinical decision-making and cohort construction for clinical AD trials. The estimated disease severity could also be used to identify transcriptomic biomarkers associated with disease progression.

In conclusion, this study provides an unsupervised method, which uses a continuous response model to estimate a disease-related latent trait. Clinically diagnosed subjects from ADNI were used to identify the correlation between observed atrophy pattern and the estimated disease severity. In the future, this model can be expanded by integrating other types of markers for estimating the disease-related latent trait. In the mean time, this model can be expanded to other types of neurodegenerative diseases using the corresponding pathologically proven cases. Such multiple-disease model would be closer to reality since mixed pathology, such as Lewy bodies or vascular disease, have been found in autopsy among many clinically diagnosed AD patients (Pao *et al.*, 2011). Such a model can reveal the underlying pathologic changes of each single patient and could be helpful for personalized treatment.

6. General Discussion

6.1 Findings of the projects

6.1.1 Summary of biomarker interaction

In this review, we presented evidences that neuroimaging (structural, molecular, diffusion-weighted and functional neuroimaging) can be used as endophenotypes to identify biomarkers that change anatomical and functional dysfunction. As a summary, pathology-related proteins have been shown their effects on brain anatomy and function: cortical A β 42 burden modulate functions in regions involved in early AD; CSF pTau/A β 42 ratio correlates with diffusivity differences and brain atrophy rate; coexist pathological marker TDP-43 showed correlations with atrophy and cognitive impairment. Sporadic AD-related genetic variant, APOE, have demonstrated its association with volumetric differences in infants and elderly people. Normal aging-related changes have been observed in regions affected at early stage of AD.

6.1.2 Summary of identification of ASR-AD

A classifier based on topographical markers extracted from clinically and pathologically proven subjects was built using a support vector machine algorithm. The classifier had a high accuracy and is comparable to a previous study published in 2008 (Klöppel *et al.*, 2008). This classifier was used to predict pathology. The classifier predicted pathology from cognitively normal people, namely CN_P+ (cognitive normal with positive pathology) subjects individually in two independent cohorts. This group of individuals was shown to be asymptomatic at risk of clinical AD (ASR-AD). Secondly, two groups, AD_P+ and AD_P-, were detected from clinically diagnosed AD patients. These groups were shown to have differences in atrophy patterns, probability of clinical conversion in follow-up study, CSF proteomic profiles and cognitive performance.

6.1.3 Summary of disease severity

To estimate a disease-related latent trait and the corresponding loadings, factor analysis model was applied to grey matter regional volumes. The expectation maximization algorithm was used to estimate the parameters of the factor analysis model. The first latent trait estimated from ADNI and 3C cohorts showed significant association with the longitudinal clinical conversion. The corresponding loadings estimated from ADNI were then used to estimate the latent trait in 3C cohort which also showed a significant association with clinical conversion. Within ADNI study, the latent trait showed significant correlations with the CSF proteomic profiles and the cognitive performance. In addition, the expected regional changes have been estimated across the whole disease spectrum.

6.2 Causality inference

In these two projects, MRI-based prediction and quantification have been constructed to provide objective measurements about Alzheimer's disease for clinical practice. Although the MRI-based predictions have shown significant associations with clinical conversion, it needs more evidence to claim that the atrophy causes clinical Alzheimer's disease. When two items are associated, three relations can be true: item A causes item B, vice versa or both of them are caused by a third item (confounding variable) independently.

This causal relationship between atrophy and clinical Alzheimer's disease still needs to be supported by other evidence. In this thesis, Hill's criteria were used to examine the causal relation. Among the conditions of the Hill's criteria, the most important one is the temporal relationship. To conclude that item A causes item B, A has to have happened before B.

In both studies, we observed that those cognitive normal subjects who have specific atrophy patterns (CN_P+ group from ASR-AD project or subjects with higher level of the estimated disease severity) are much more likely to develop clinical Alzheimer's disease. These observations support our hypothesis.

On the other hand, in both projects, we observed cognitive normal subjects predicted to have no pathology (CN_P-) or with low level of the estimated disease severity converted to clinical MCI or AD. One interpretation can be that those subjects had already the disease but not severe enough to be classified as CN_P+ or with a higher value of the estimated severity. Atrophy progressed in those subjects over time. Symptoms appeared when the stage of the destruction of the inferior temporal and lateral temporal cortex reached. Another explanation is that those subjects suffered from other neurodegenerative diseases therefore they did not have typical AD-pathology related atrophy. But the symptoms were similar to AD and since clinical diagnostic criteria cannot discriminate Alzheimer's disease from other types of neurodegenerative disease (Beach et al., 2012), they were diagnosed as AD in the follow-up study.

In contrast to CN_P- subjects converting to AD, there are subjects predicted to have pathology-related atrophy (CN_P+), who have not converted to clinical AD. The possible reasons are : previous studies have shown factors, such as education, which can reduce the influence of pathology on the clinical expression of symptoms (Brayne *et al.*, 2010). Other factors, like lifestyle, can slow down the progress of symptoms (Khalsa, 2015).

	Identification of ASR-AD	Estimation of disease severity
Temporal Relationship	Clinical conversion to MCI (or AD) in follow-up of CN_P+ and CN_P- were examined.	Clinical conversion in follow-up of cognitive normal (to MCI or AD) and MCI groups (to AD) were examined.
Strength	CN_P+ had significantly higher risk (hazard ratio) than CN_P- to convert to MCI or clinical AD.	In cognitive normal / MCI group, subjects with higher values in the estimated severity were more likely to proceed to clinical diagnosis.
Dose-response		The estimated severity correlated with cognitive performance and clinical diagnosis.
Consistency	1) The results from the two disease models (chapter 4 and 5) were correlated, see in Supplementary Figure 7; 2) both results have been replicated using an independent cohort	
Plausibility	Results from the two disease models correlated with pathological markers.	
Consideration of Alternate Explanations (Analogy)	APOE, SNPs, age, gender (adjusted confounders)	APOE, age, gender, CSF proteins (adjusted confounders)
Coherence	Cognitive performance correlates atrophy patterns have been found in several studies (Braak and Braak, 1995; Ossenkoppele, Schonhaut, <i>et al.</i> , 2015).	
Experiment	This criterion is not applied, due to the fact that there is no cure to reduce the level of exposure.	
Specificity	This criterion is not applied. Outcomes are likely to have multiple factors influencing them. It is highly unlikely that we will find a one-to-one cause-effect relationship in sporadic AD.	

Table 9 Summary of two experiments results for testing a causality relation between atrophy and clinical AD conversion

Each of Hill's criteria was used to test a causality relation between atrophy and clinical AD conversion.

Apart from temporary relationship, other criteria were tested as well (summarized in Table 9). Due to the characteristics of Alzheimer's disease, only a subset of criteria can be (partially) tested. Several studies have been published to provide evidence between brain volumetric change and its relation with memory deficits as mentioned in chapter 0 (coherence criterion). According to the theory of A β 42 and Tau, it is more accepted that A β 42 and/or Tau trigger the neuron death. We demonstrated the relation

between the results from our disease models and these pathology markers to test the plausibility. For consistency, we applied different methods in these two projects and the predicted results are significantly correlated (shown in Supplementary Figure 7). In both projects, results could be consistently replicated in different cohorts collected among different populations (a research cohort ADNI from America, an epidemiological cohort 3C from France). In the project of estimating disease severity two criteria, strength and dose-response, have been demonstrated by comparing the estimated severity with clinical diagnosis and cognitive measurements. These criteria could not be shown in the project of identification of ASR-AD, because clinical measurements within cognitive normal subjects have small variances.

Some of the criteria cannot be tested with current technology. As Alzheimer's disease is irreversible, the Hill's experiment evidence criterion cannot be satisfied. The Hill's specificity criterion cannot be tested due to the fact that sporadic Alzheimer's disease is a complex disease and it is likely to have multiple factors causing the clinical symptoms, rather than a one-to-one cause-effect relationship.

6.3 Methodological discussion

6.3.1 Normalization within feature or within subject

Throughout both projects, data from multiple sources have been combined, different cohorts and studies. The estimated grey matter volumes can be influenced by several data recruiting factors, especially when data are collected from multisite research centres, using different scanner types, MR strength and imaging protocols. These factors introduce bias by affecting the intensity of the images. Apart from recruiting bias, imaging preprocessing steps, such as Dartel normalization, can also introduce artefacts to voxels near meninges and white matter.

Using "within feature scaling", as suggested by SVM package (Chang and Lin, 2011), each of the voxels needs to be scaled to the range [0, 1]. Due to the scaling, voxels near the boundaries are attributed within the same range as disease-related informative voxels, thus amplifying the noise. The results showed that voxels near meninges in the frontal lobe were significantly different between the predicted groups. A second problem with within feature scaling is its sensitivity to outliers: removing one subject can bring significant change to the prediction labels.

Since many factors that can influence the estimated grey matter volume (Marchewka *et al.*, 2014), such as scanner types and field strength, the estimated data cannot be compared directly between subjects. Normalization within subject is used to set the estimated volumes of all voxels from one subject to a mean of 0 and a standard deviation of 1. For example, in training AD subjects, voxel A from frontal lobe has a larger volume than voxel B from temporal lobe. But this relation is not

observed in healthy group. New subjects having this pattern will be classified as AD patients. So within subject scaling is robust to artifacts and outliers.

Another type of normalization calculates the ratio of each region or voxel to the total estimated grey matter volume. By doing so, a regional volume (in litter) can be replaced by the proportion of this region to a whole. However, this step requires that the volume be transformed into an unlimited continuous variable. Many models, such as linear regression, assumed that the response should be a continuous unbounded variable.

These normalization steps are very important as different cohorts are integrated in a single project for the purposes of increasing statistical power and replicating the results. By using normalization methods, the biases due to multi-centers and study protocols can be reduced.

6.3.2 Correction for covariates' effects

As previously stated, bias due to study centers can be reduced by applying normalization methods. However, bias can be introduced between subjects by other factors, such as age, TIV and gender. Most studies use age, TIV and gender as covariates, to identify regions with significantly different volumes between cases and controls (Mori *et al.*, 2014). However, many studies do not correct volumes before building a classifier (Klöppel *et al.*, 2008). To identify ASR-AD subjects, the data used in this study was corrected for the effects of age and TIV by using linear regression. The effects of age and TIV were assumed to be linear, since all subjects were above 50 years. The results showed no aging bias in the prediction of ASR-AD in ADNI subjects, but a minor effect remained in the 3C cohorts.

A future study could look at other situations, such as brain changes across lifespan. Brain regions grow rapidly at early childhood and decrease slowly from around 40 years old. A quadratic age effect would be a better choice to fit the change of brain volume.

6.3.3 Feature dimension or subject dimension

In machine learning, high dimensionality tends to bring an over-fitting problem during the classifier training procedure. This was avoided by building a dot product kernel and using it to train a classifier. This dot product kernel represents the similarities between subjects. When the support vector machine algorithm was applied on the similarity matrix, a hyperplane was obtained which can best separate the cases from controls in the subject similarity space.

6.3.4 PCA or ML estimation

The factor analysis model can be estimated by a PCA-based method or by expectation maximization (EM). Both methods estimate the factor loadings resulting in the covariance of the observed variables,

which are the grey matter regional volumes. The PCA-based method relies on the eigen-decomposition or the singular value decomposition to construct eigenvectors along which the data set has the maximum variance. The EM-based method finds the factor loadings which maximize the data likelihood. The results have suggested that the estimated factors based on the combination of the two methods are more correlated to clinical diagnosis. In this project, the PCA-based method was applied to initialize the factor loadings and then maximize the data likelihood by an iterative procedure.

6.3.5 Common genetic variants

In the last decade several studies have focused on GWAS to identify the genetic risk factors for neuronal disorder diseases, such as mood disorders and psychotic disorders. However, these neuronal disorder diseases are mainly diagnosed based on symptoms. The GWAS method relies on precise phenotypes which should be not replaced by symptoms. Several common genetic variants, SNPs, identified in the neuronal disorder diseases could not be replicated by other studies (Freitag, 2007). Most neuronal disorder diseases, including Alzheimer's, are complex and multigenic. In the project of identification of ASR-AD, 16 SNPs were identified from previous studies have been tested individually for their associations on longitudinal clinical conversions. These results agreed with the findings of previous studies. Genetic risk score (GRS) was proposed to test a list of SNPs' effects when using a relatively small number of subjects, by combining several SNPs together. In the project of identifying ASR-AD, a weighted GRS (wGRS) was built based on 16 SNPs. In addition to this wGRS, another wGRS was calculated using the 16 SNPs and the APOE genotype. This new wGRS demonstrated a significant association with longitudinal clinical conversion, suggesting that the wGRS's effect was driven by APOE.

In addition to the 16 SNPs, one SNP (rs3785883) from the MAPT gene has been shown to have effects on longitudinal memory decline and disease onset (Cruchaga *et al.*, 2010; Kauwe *et al.*, 2008; Peterson *et al.*, 2014). In order to identify this gene's effect on grey matter rate of change, two years of T1-weighted MRI scans were collected for subjects carrying different rs3785883 genotypes. All scans went through standard preprocessing steps using SPM12. No region showed a significant difference in the rate of change of volume that is associated with MAPT genotypes, after adjustment of age, gender and APOE- ϵ 4. Two cohorts, ADNI and 3C, were used and results were consistent.

These results could be due to the fact that the number of subjects, 23 were carrying two risk alleles and 203 were carrying one risk allele (out of 716), was small or that the effect on memory decline is not related to grey matter volumetric change but occurs via another pathway, such as iron accumulation or connectome properties.

6.4 The extension of the three principles

Structural neuroimaging is easily performed and less expensive than molecular imaging. In previous chapters, the usage of T1-weighted neuroimaging in Alzheimer's disease has been shown. However building a classifier for each type of neurodegenerative disease is not practical in clinical routine. Although several neurodegenerative diseases have specific pathology patterns, such as neurofibrillary tangles, amyloid plaques in Alzheimer's disease and Lewy bodies in Parkinson's disease and Lewy body dementia, comorbidity is common in patients, as most of the cases have a mixture of different types of pathophysiologies (Harper *et al.*, 2014). Progressive neuronal loss appears in all types of neurodegenerative disease with specific atrophy patterns. However, in many of these diseases, overlapped atrophy patterns at certain stages can be observed. Such characteristics of brain structure and molecular markers suggest a usage of multiclass classification or clustering in neuroimaging and molecular measurements. As an extension to the challenges mentioned in section 1.2, this section will discuss the optional algorithms, methods and new technologies that can be useful in clinical practice and the discovery of the underlying mechanisms.

6.4.1 Principle 1: computational anatomy

Different modalities for studying neurodegenerative disease

To identify the cause of neuronal loss in neurodegenerative disease, several types of molecular markers are worth exploring. Enzymes and proteins are essential in forming the components of neurons and glia and involved in cellular functions. As mentioned in the introduction chapter, in several neurodegenerative diseases, soluble protein oligomers have been found to confer synaptic structure and plasticity (Querfurth and LaFerla, 2010). Therefore, disease-related protein structural features and sequential changes can be identified using brain morphological patterns as endophenotypes. Such change might include quantitative (levels) and qualitative (structures) changes. Based on the identified disease related proteins, the corresponding genetic transcription and translation pathways can be revealed. At this level, both genetic and epigenetic markers should be studied.

Calcium and iron are essential but toxic at high levels to neurons (Singh *et al.*, 2014). The iron levels can be directly measured using structural neuroimaging techniques (R2 star). Iron load is one of the most developed markers in Parkinson's disease, as it accumulates in the substantia nigra and the cortex (Reeve *et al.*, 2014). Whether such accumulation exists in other types of neurodegenerative disease and its temporal relation with neuronal loss and clinical conversion still need to be examined.

New technologies, like tau PET, are emerging and could be useful for research. F-18 T807 (¹⁸F-AV-1451) is the best developed tau pathology PET tracer and can be potentially used as a diagnostic tool *in vivo*. F-18 T807 has been found to strongly associate with tau neuropathology in MAPT mutation

carriers (Smith *et al.*, 2016). Since MAPT mutation causes one subtype of frontotemporal lobe dementia (Van Der Zee and Broeckhoven, 2014), the propagation of tau protein can be compared with the neuronal loss in the corresponding regions. Such comparisons can be performed through all stages of the disease and therefore a temporal relationship between tau pathology and atrophy and memory decline can be determined.

6.4.2 Principle 2: Disease modeling

Multiclass classification and semi-supervised clustering

When manifold pathologically proven data is available, classification algorithms, like support vector machine or logistic regression, can be used to build a multiclass classifier. Algorithms, such as logistic regression, provide probabilities of each class when assigning a new data point. The application of this type of classifier can help clinicians to make a diagnosis with a measure of the reliability. Such classifiers can be constructed and tested using different types of modalities of data, like neuroimaging and molecular markers or a combination of them.

In reality, pathology proven data is rare. Most studies about neurodegenerative disease rely on clinical diagnosis. As explained above, the clinical diagnoses are not precise. Semi-supervised clustering may be useful to solve this problem. Unlike unsupervised clustering algorithms, semi-supervised methods use labeled data points (pathology proven data) as seeds to build the initial prototypes and / or set the constraints. With the initial probability distributions of all groups, then algorithms like K-means can be applied to iteratively assign the unlabelled data, clinical diagnosed subjects. Mismatched labels from clinical and clustering can be obtained due to the fact that many diseases have similar symptoms.

Applying a semi-supervised clustering algorithm on all types of neurodegenerative disease, can reveal the structure between different diseases and subpopulations. In addition, a clustering of all diseases based on brain structures and molecular markers can be compared. For example, Lewy bodies, as pathological markers which are mainly composed of α -synuclein, exist in multiple neurodegenerative diseases, such as Lewy body dementia, Parkinson disease and multiple system atrophy. Recent studies have shown that subgroups of Parkinson's brains have different α -synuclein structural characteristics (Bousset *et al.*, 2013; Guo *et al.*, 2013). Based on such findings, one hypothesis could be that diseases can be differentiated by using different modalities of data such as brain structural features and protein sequences and structural information. Such comparisons may identify subpopulations, like the three subtypes of Alzheimer's disease (Whitwell *et al.*, 2012).

Applying a clustering algorithm to all types of neurodegenerative disease requires a consideration about the usage of the method. For example, the PCA method requires that the principal components

are orthogonal, which means all diseases are independent. Semi-supervised learning algorithm makes the assumption that the unlabeled data are carrying the information related to the disease.

When performing a multiclass classification or clustering, matched aging healthy subjects should be involved for the following reason. As explained in previous sections, normal aging is the main confounder in neurodegenerative diseases. Therefore a group of healthy subjects is necessary to model the normal aging effects.

Modular approach for integrative analysis

Investigating one type of data set at a time may miss potential interactions between multiple types of measurement. For a complex disease, like Alzheimer's disease, clinical diagnosed groups are not homogeneous as explained in previous chapters. Identify those clusters or co-modules can reduce the complexity of studying the disease mechanism. Such modules should have coherent patterns across two data sets from the same subjects. In one of my collaborative study, we used rates of regional atrophy and gene expression to identify co-modules using a Ping-Pong algorithm (Kutalik *et al.*, 2008). Through this experiment, a set of genes showed significant association with longitudinal brain atrophy. Moreover, the expression of this set of genes showed significant association with longitudinal clinical conversion among cognitive normal and MCI group. Their roles in neurodegenerative are still under investigation.

6.4.3 Principle 3: Biomarkers interaction used for construction and validation models

Disease prevention factors

Several factors have been found to protect against the development of Alzheimer's disease. As mentioned in previous sections, high education levels have shown to reduce the impact of pathology on the symptoms' expression. Other studies have shown that healthy diet and sport activities have protective roles on cardiovascular factors which significantly associate with Alzheimer's disease (Pase *et al.*, 2016). The mechanisms of these factors might be involving the change of brain structure, such as preventing the shrinkage of the neurons, increasing the plasticity of the brain or the compensation by other regions. These questions still need to be addressed in future studies and be useful in understanding the brain and preventing disease in the general population.

Coexisting symptoms

Some symptoms related to depression can be observed among a subset of Alzheimer's disease patients (Steinberg *et al.*, 2004). It has been shown that bipolar disorder patients have progressive loss in hippocampal and fusiform gyri (Moorhead *et al.*, 2007), which are important regions significantly

affected by Alzheimer's disease. Bipolar disorder can increase the risk of dementia onset (Almeida *et al.*, 2016). It is still unclear what the link is between bipolar disorder and Alzheimer's disease, for instance whether they share the same biological mechanism. Future studies can focus on the identification of pathways of these two diseases and compare the similarities and differences between them.

Another common co-existing symptom in Alzheimer's disease is sleep disorder, including sleep-wake cycle and circadian rhythms (Musiek *et al.*, 2015). It has been reported that sleep can increase the volume of extracellular fluid in the brain and therefore metabolites and proteins can be "washed out" (Xie *et al.*, 2013). The proposed functions of sleep in Alzheimer's disease involve reducing the production and enhancing the clearance of A β 42 (Musiek *et al.*, 2015). Therefore, sleep disorder is considered to conceivably facilitate the formation of amyloid plaques, from small plaques to large plaques (Musiek *et al.*, 2015). Future research is needed to clarify the function of sleep in "transporting" the propagated proteins and this understanding might be helpful to develop a new treatment strategy.

7. Reference

3C Study Group. Vascular factors and risk of dementia: Design of the Three-City Study and baseline characteristics of the study population. *Neuroepidemiology* 2003; 22: 316–325.

Adriaanse SM, Sanz-Arigita EJ, Binnewijzend MAA, Ossenkoppele R, Tolboom N, van Assema DME, et al. Amyloid and Its Association with Default Network Integrity in Alzheimer's Disease. *Hum. Brain Mapp.* 2014; 35: 779–791.

Ahmed M, Davis J, Aucoin D, Sato T, Ahuja S, Aimoto S, et al. Structural conversion of neurotoxic amyloid-beta(1-42) oligomers to fibrils. *Nat. Struct. Mol. Biol.* 2010; 17: 561–7.

Alexander-Bloch A, Giedd JN, Bullmore E. Imaging structural co-variance between human brain regions *Nat Rev Neurosci* 2013; 14: 322–336.

Almeida OP, McCaul K, Hankey GJ, Yeap BB, Golledge J, Flicker L, et al. Risk of dementia and death in community-dwelling older men with bipolar disorder. *Br. J. Psychiatry* 2016; 209: 121–6.

Alpérovitch A, Amouyel P, Dartigues JF, Ducimetière P, Mazoyer B, Ritchie K, et al. Epidemiological studies on ageing in France: From PAQUID to the Three-City study. *Comptes Rendus - Biol.* 2002; 325: 665–672.

Anderson C a, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. Data quality control in genetic case-control association studies *Nat. Protoc.* 2010; 5: 1564–1573.

Ashburner J. Computational anatomy with the SPM software. *Magn. Reson. Imaging* 2009; 27: 1163–1174.

Baker FB. *The Basics of Item Response Theory* 2001.

Beach TG, Monsell SE, Kukull W a. Accuracy of the clinical diagnosis of alzheimer disease at national institute on aging alzheimer's disease centers, 2005-2010. *J Neuropathol Exp Neurol* 2012; 71: 266–273.

Belleville S, Fouquet C, Duchesne S, Collins DL, Hudon C. Detecting early preclinical Alzheimer's disease via cognition, neuropsychiatry, and neuroimaging: qualitative review and recommendations for testing. *J. Alzheimer's Dis.* 2014; 42 Suppl 4: S375-82.

Bousset L, Pieri L, Ruiz-Arlandis G, Gath J, Jensen PH, Habenstein B, et al. Structural and functional characterization of two alpha-synuclein strains. *Nat. Commun.* 2013; 4: 2575.

Braak H, Braak E. Staging of Alzheimer's disease-related neurofibrillary changes. *Neurobiol. Aging* 1995; 16: 271–278.

Brayne C, Ince PG, Keage HAD, McKeith IG, Matthews FE, Polvikoski T, et al. Education, the brain and dementia: Neuroprotection or compensation? *Brain* 2010; 133: 2210–2216.

Brett M, Penny W, Kiebel S. 14. Introduction to Random Field Theory. In: *Human Brain Function*. 2003. p. 1–23.

Bron EE, Steketee RME, Houston GC, Oliver RA, Achterberg HC, Loog M, et al. Diagnostic

Classification of Arterial Spin Labeling and Structural MRI in Presenile Early Stage Dementia. *Hum. Brain Mapp.* 2014; 35: 4916–4931.

Buchhave P, Minthon L, Zetterberg H, Wallin a. K, Blennow K, Hansson O. Cerebrospinal Fluid Levels of β -amyloid 1-42, but Not of Tau, Are Fully Changed Already 5 to 10 Years Before the Onset of Alzheimer Dementia. *Arch. Gen. Psychiatry* 2012; 69: 98–106.

Buerger K, Alafuzoff I, Ewers M, Pirttilä T, Zinkowski R, Hampel H. No correlation between CSF tau protein phosphorylated at threonine 181 with neocortical neurofibrillary pathology in Alzheimer's disease [2]. *Brain* 2007; 130: 180–181.

Buerger K, Ewers M, Pirttilä T, Zinkowski R, Alafuzoff I, Teipel SJ, et al. CSF phosphorylated tau protein correlates with neocortical neurofibrillary pathology in Alzheimer's disease. *Brain* 2006; 129: 3035–3041.

Bush WS, Moore JH. Chapter 11: Genome-Wide Association Studies. *PLoS Comput. Biol.* 2012; 8

Carrasquillo MM, Crook JE, Pedraza O, Thomas CS, Pankratz VS, Allen M, et al. Late-onset Alzheimer's risk variants in memory decline, incident mild cognitive impairment, and Alzheimer's disease. *Neurobiol. Aging* 2015; 36: 60–7.

Chang C-C, Lin C-J. LIBSVM: A Library for Support Vector Machines *ACM Trans. Intell. Syst. Technol.* 2011; 2: 27:1--27:27.

Chen H, Poon A, Yeung C, Helms C, Pons J, Bowcock AM, et al. A genetic risk score combining ten psoriasis risk loci improves disease prediction. *PLoS One* 2011; 6: 1–7.

Chételat G, Landeau B, Eustache F, Mézenge F, Viader F, De La Sayette V, et al. Using voxel-based morphometry to map the structural changes associated with rapid conversion in MCI: A longitudinal MRI study. *Neuroimage* 2005; 27: 934–946.

Clarke GM, Anderson C a, Pettersson FH, Cardon LR, Andrew P. Basic statistical analysis in genetic case-control studies. *Nat. Protoc.* 2011; 6: 121–133.

Coupé P, Fonov VS, Bernard C, Zandifar A, Eskildsen SF, Helmer C, et al. Detection of Alzheimer's disease signature in MR images seven years before conversion to dementia: Toward an early individual prognosis. *Hum. Brain Mapp.* 2015; 4770: 4758–4770.

Cox DG, Kraft P. Quantification of the power of Hardy-Weinberg equilibrium testing to detect genotyping error. *Hum. Hered.* 2006; 61: 10–14.

Cripps D, Thomas SN, Jeng Y, Yang F, Davies P, Yang AJ. Alzheimer disease-specific conformation of hyperphosphorylated paired helical filament-Tau is polyubiquitinated through Lys-48, Lys-11, and Lys-6 ubiquitin conjugation. *J. Biol. Chem.* 2006; 281: 10825–10838.

Cruchaga C, Kauwe JSK, Mayo K, Spiegel N, Cruchaga C, Kauwe JSK, et al. SNPs associated with cerebrospinal fluid Phospho-tau levels influence rate of decline in alzheimer's disease. *PLoS Genet.* 2010; 6: 1–10.

Cui J, Zufferey V, Kherif F. In-vivo brain neuroimaging provides a gateway for integrating

biological and clinical biomarkers of Alzheimer's disease *Curr. Opin. Neurol.* 2015; 28: 351–357.

Davatzikos C, Xu F, An Y, Fan Y, Resnick SM. Longitudinal progression of Alzheimers-like patterns of atrophy in normal older adults: The SPARE-AD index. *Brain* 2009; 132: 2026–2035.

Dean DC, Jerskey B a, Chen K, Protas H, Thiyyagura P, Roontiva A, et al. Brain differences in infants at differential genetic risk for late-onset Alzheimer disease: a cross-sectional imaging study. *JAMA Neurol.* 2014; 71: 11–22.

Desikan RS, Thompson WK, Holland D, Hess CP, Brewer JB, Zetterberg H, et al. The Role of Clusterin in Amyloid-beta-Associated Neurodegeneration *JAMA Neurol* 2014; 71: 180–187.

Doraiswamy PM, Sperling R a, Johnson K, Reiman EM, Wong TZ, Sabbagh MN, et al. Florbetapir F 18 amyloid PET and 36-month cognitive decline:a prospective multicenter study. *Mol. Psychiatry* 2014: 1–8.

Draganski B, Kherif F, Lutti A. Computational anatomy for studying use-dependant brain plasticity. *Front. Hum. Neurosci.* 2014; 8: 380.

Drzezga a, Grimmer T, Henriksen G. volume in Alzheimer disease Effect of APOE genotype on amyloid plaque load and gray matter volume in Alzheimer disease. 2009

Dubois B, Feldman HH, Jacova C, DeKosky ST, Barberger-Gateau P, Cummings J, et al. Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. *Lancet Neurol.* 2007; 6: 734–746.

Dubois B, Feldman HH, Jacova C, Hampel H, Molinuevo JL, Blennow K, et al. Advancing research diagnostic criteria for Alzheimer's disease: the IWG-2 criteria. *Lancet Neurol.* 2014; 13: 614–29.

Dubois B, Hampel H, Feldman HH, Scheltens P, Aisen P, Andrieu S, et al. Preclinical Alzheimer's disease: Definition, natural history, and diagnostic criteria 2016.

Fagan AM, Roe CM, Xiong C, Mintun M a, Morris JC, Holtzman DM. Cerebrospinal fluid tau/beta-amyloid(42) ratio as a prediction of cognitive decline in nondemented older adults. *Arch. Neurol.* 2007; 64: 343–349.

Ferrando PJ. Theoretical and Empirical Comparisons between Two Models for Continuous Item Response. *Multivariate Behav. Res.* 2002; 37: 521–542.

Fleisher AS, Chen K, Quiroz YT, Jakimovich LJ, Gomez MG, Langois CM, et al. Florbetapir PET analysis of amyloid- β deposition in the presenilin 1 E280A autosomal dominant Alzheimer's disease kindred: a cross-sectional study. *Lancet. Neurol.* 2012; 11: 1057–65.

Forde NJ, Ronan L, Suckling J, Scanlon C, Neary S, Holleran L, et al. Structural neuroimaging correlates of allelic variation of the BDNF val66met polymorphism. *Neuroimage* 2014; 90: 280–9.

Fox J, Weisberg S. *An R Companion to Applied Regression* Sage 2011

Freitag CM. The genetics of autistic disorders and its clinical relevance: a review of the literature. *Mol Psychiatry* 2007; 12: 2–22.

Frisoni GB, Fox NC, Jack CR, Scheltens P, Thompson PM. The clinical use of structural MRI in Alzheimer disease. *Nat. Rev. Neurol.* 2010; 6: 67–77.

Ghahramani Z, Hinton GE. The EM Algorithm for Mixtures of Factor Analyzers. *Compute* 1997: 1–8.

Giannakopoulos P, Herrmann FR, Bussi re T, Bouras C, Kovari E, Perl DP, et al. Tangle and neuron numbers, but not amyloid load, predict cognitive status in Alzheimer’s disease. *Neurology* 2003; 60: 1495–500.

Glymour MM, Ch ne G, Tzourio C, Dufouil C. Brain MRI markers and dropout in a longitudinal study of cognitive aging: The Three-City Dijon Study. *Neurology* 2012; 79: 1340–1348.

Goedert M, Jakes R. Mutations causing neurodegenerative tauopathies. *Biochim. Biophys. Acta - Mol. Basis Dis.* 2005; 1739: 240–250.

Guo JL, Covell DJ, Daniels JP, Iba M, Stieber A, Zhang B, et al. Distinct α -synuclein strains differentially promote tau inclusions in neurons *Cell* 2013; 154: 103–117.

Hafler DA, Jager PL De. Applying a new generation of genetic maps to understand human inflammatory disease *Nat. Rev. Immunol.* 2005; 5: 83–91.

Hardy J, Selkoe DJ. The amyloid hypothesis of Alzheimer’s disease: progress and problems on the road to therapeutics. *Science* 2002; 297: 353–356.

Harold D, Abraham R, Hollingworth P, Sims R, Gerrish A, Hamshere ML, et al. Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer’s disease. *Nat. Genet.* 2009; 41: 1088–1093.

Harper L, Barkhof F, Scheltens P, Schott JM, Fox NC. An algorithmic approach to structural imaging in dementia. *J. Neurol. Neurosurg. Psychiatry* 2014; 85: 692–8.

Van Harten AC, Visser PJ, Pijnenburg Y a L, Teunissen CE, Blankenstein M a., Scheltens P, et al. Cerebrospinal fluid AB42 is the best predictor of clinical progression in patients with subjective complaints *Alzheimer’s Dement.* 2013; 9: 481–487.

Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning.* Elements 2009; 1: 337–387.

Hohman TJ, Koran MEI, Thornton-Wells TA. Interactions between GSK3 β and amyloid genes explain variance in amyloid burden. *Neurobiol. Aging* 2014; 35: 460–5.

Hollingworth P, Harold D, Sims R, Gerrish A, Lambert J-C, Carrasquillo MM, et al. Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer’s disease. *Nat. Genet.* 2011; 43: 429–435.

H ltt  M, Hansson O, Andreasson U, Hertze J, Minthon L, N gga K, et al. Evaluating Amyloid- β Oligomers in Cerebrospinal Fluid as a Biomarker for Alzheimer’s Disease. *PLoS*

One 2013; 8

Huijbers W, Mormino EC, Wigman SE, Ward a. M, Vannini P, McLaren DG, et al. Amyloid Deposition Is Linked to Aberrant Entorhinal Activity among Cognitively Normal Older Adults *J. Neurosci.* 2014; 34: 5200–5210.

Jack CR, Bennett DA, Blennow K, Carrillo MC, Feldman HH, Frisoni GB, et al. A/T/N: An unbiased descriptive classification scheme for Alzheimer disease biomarkers. *Neurology* 2016; 87: 539–547.

Jack CR, Holtzman DM. Biomarker modeling of alzheimer's disease. *Neuron* 2013; 80: 1347–1358.

Jack CR, 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.

Jack CR, Wiste HJ, Weigand SD, Rocca WA, Knopman DS, Mielke MM, et al. Age-specific population frequencies of cerebral β -amyloidosis and neurodegeneration among people with normal cognitive function aged 50–89 years: a cross-sectional study. *Lancet Neurol.* 2014; 13: 997–1005.

Jansen WJ, Ossenkoppele R, Knol DL, Tijms BM, Scheltens P, Verhey FRJ, et al. Prevalence of Cerebral Amyloid Pathology in Persons Without Dementia *JAMA* 2015; 313: 1924.

Jiang J, Sachdev P, Lipnicki DM, Zhang H, Liu T, Zhu W, et al. A longitudinal study of brain atrophy over two years in community-dwelling older individuals *Neuroimage* 2014; 86: 203–211.

Josephs K a., Whitwell JL, Weigand SD, Murray ME, Tosakulwong N, Liesinger AM, et al. TDP-43 is a key player in the clinical features associated with Alzheimer's disease. *Acta Neuropathol.* 2014; 127: 811–824.

Jucker M, Walker LC. Self-propagation of pathogenic protein aggregates in neurodegenerative diseases. *Nature* 2013; 501: 45–51.

Karas GB, Scheltens P, Rombouts SARB, Visser PJ, Van Schijndel RA, Fox NC, et al. Global and local gray matter loss in mild cognitive impairment and Alzheimer's disease. *Neuroimage* 2004; 23: 708–716.

Kauwe JSK, Cruchaga C, Mayo K, Fenoglio C, Bertelsen S, Nowotny P, et al. Variation in MAPT is associated with cerebrospinal fluid tau levels in the presence of amyloid-beta deposition. *Proc. Natl. Acad. Sci. U. S. A.* 2008; 105: 8050–8054.

Khalsa DS. Stress, meditation, and Alzheimer's disease prevention: Where the evidence stands. *J. Alzheimer's Dis.* 2015; 48: 1–12.

Khan U a, Liu L, Provenzano F a, Berman DE, Profaci CP, Sloan R, et al. Molecular drivers and cortical spread of lateral entorhinal cortex dysfunction in preclinical Alzheimer's disease. *Nat. Neurosci.* 2014; 17: 304–11.

Kherif F, Josse G, Seghier ML, Price CJ. The main sources of intersubject variability in

neuronal activation for reading aloud. *J. Cogn. Neurosci.* 2009; 21: 654–668.

Klöppel S, Stonnington CM, Chu C, Draganski B, Scahill RI, Rohrer JD, et al. Automatic classification of MR scans in Alzheimer's disease. *Brain* 2008; 131: 681–689.

Koran MEI, Hohman TJ, Meda SA, Thornton-Wells TA. Genetic interactions within inositol-related pathways are associated with longitudinal changes in ventricle size. *J. Alzheimer's Dis.* 2014; 38: 145–154.

Koran MEI, Hohman TJ, Thornton-Wells T a. Genetic interactions found between calcium channel genes modulate amyloid load measured by positron emission tomography. *Hum. Genet.* 2014; 133: 85–93.

Krasuski JS, Alexander GE, Horwitz B, Daly EM, Murphy DG, Rapoport SI, et al. Volumes of medial temporal lobe structures in patients with Alzheimer's disease and mild cognitive impairment (and in healthy controls). *Biol Psychiatry* 1998; 43: 60–68.

Kutalik Z, Beckmann JS, Bergmann S. A modular approach for integrative analysis of large-scale gene-expression and drug-response data. *Nat. Biotechnol.* 2008; 26: 531–9.

Lambert J-C, Heath S, Even G, Campion D, Sleegers K, Hiltunen M, et al. Genome-wide association study identifies variants at *CLU* and *CR1* associated with Alzheimer's disease. *Nat. Genet.* 2009; 41: 1094–1099.

Lambert JC, Ibrahim-Verbaas C a, Harold D, Naj a C, Sims R, Bellenguez C, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat. Genet.* 2013; 45: 1452–8.

Lehmann M, Ghosh PM, Madison C, Karydas A, Coppola G, O'Neil JP, et al. Greater medial temporal hypometabolism and lower cortical amyloid burden in ApoE4-positive AD patients. *Neurol. Neurosurgery, Psychiatry* 2014; 85: 266–73.

Lesaffre E, Rizopoulos D, Tsonaka R. The logistic transform for bounded outcome scores. *Biostatistics* 2007; 8: 72–85.

Lillemark L, Sørensen L, Pai A, Dam EB, Nielsen M. Brain region's relative proximity as marker for Alzheimer's disease based on structural MRI. *BMC Med. Imaging* 2014; 14: 21.

Lim J-S, Park YH, Jang J-W, Park SY, Kim S. Differential White Matter Connectivity in Early Mild Cognitive Impairment According to CSF Biomarkers. *PLoS One* 2014; 9: e91400.

Lim YY, Villemagne VL, Laws SM, Ames D, Pietrzak RH, Ellis K a., et al. Effect of BDNF Val66Met on memory decline and hippocampal atrophy in prodromal alzheimer's disease: A preliminary study. *PLoS One* 2014; 9: 10–14.

Ly M, Canu E, Xu G, Oh J, McLaren DG, Dowling NM, et al. Midlife measurements of white matter microstructure predict subsequent regional white matter atrophy in healthy adults. *Hum. Brain Mapp.* 2014; 35: 2044–2054.

Lyall DM, Harris SE, Bastin ME, Muñoz Maniega S, Murray C, Lutz MW, et al. Alzheimer's disease susceptibility genes *APOE* and *TOMM40*, and brain white matter integrity in the Lothian Birth Cohort 1936. *Neurobiol. Aging* 2014; 35: 1513.e25-33.

Manning EN, Barnes J, Cash DM, Bartlett JW, Leung KK, Ourselin S, et al. APOE ϵ 4 is associated with disproportionate progressive hippocampal atrophy in AD. *PLoS One* 2014; 9: 1–8.

Marchewka A, Kherif F, Krueger G, Grabowska A, Frackowiak R, Draganski B. Influence of magnetic field strength and image registration strategy on voxel-based morphometry in a study of Alzheimer's disease. *Hum. Brain Mapp.* 2014; 35: 1865–74.

Mckhann G, Drachman D, Folstein M. 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. 1984

Moorhead TWJ, McKirdy J, Sussmann JED, Hall J, Lawrie SM, Johnstone EC, et al. Progressive Gray Matter Loss in Patients with Bipolar Disorder. *Biol. Psychiatry* 2007; 62: 894–900.

Mori T, Shimada H, Shinotoh H, Hirano S, Eguchi Y, Yamada M, et al. Apathy correlates with prefrontal amyloid β deposition in Alzheimer's disease. *J. Neurol. Neurosurg. Psychiatry* 2014; 85: 449–55.

Morris JC. The Clinical Dementia Rating (CDR): current version and scoring rules. *Neurology* 1993; 43: 2412–2414.

Musiek ES, Xiong DD, Holtzman DM. Sleep, circadian rhythms, and the pathogenesis of Alzheimer disease. *Exp. Mol. Med.* 2015; 47: e148.

Naj AC, Jun G, Beecham GW, Wang L-S, Vardarajan BN, Buross J, et al. Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. *Nat. Genet.* 2011; 43: 436–441.

Newlander SM, Chu A, Sinha US, Lu PH, Bartzokis G. Methodological improvements in voxel-based analysis of diffusion tensor images: Applications to study the impact of apolipoprotein e on white matter integrity. *J. Magn. Reson. Imaging* 2014; 39: 387–397.

Olsson B, Lautner R, Andreasson U, Öhrfelt A, Portelius E, Bjerke M, et al. CSF and blood biomarkers for the diagnosis of Alzheimer's disease: A systematic review and meta-analysis. *Lancet Neurol.* 2016

Ossenkoppele R, Jansen WJ, Rabinovici GD, Knol DL, van der Flier WM, van Berckel BNM, et al. Prevalence of amyloid PET positivity in dementia syndromes: a meta-analysis. *Jama* 2015; 313: 1939–49.

Ossenkoppele R, Schonhaut DR, Baker SL, O'Neil JP, Janabi M, Ghosh PM, et al. Tau, amyloid, and hypometabolism in a patient with posterior cortical atrophy. *Ann. Neurol.* 2015; 77: 338–342.

Pao WC, Dickson DW, Crook JE, Finch N a, Rademakers R, Graff-Radford NR. Hippocampal sclerosis in the elderly: genetic and pathologic findings, some mimicking Alzheimer disease clinically. *Alzheimer Dis. Assoc. Disord.* 2011; 25: 364–8.

Pase MP, Beiser A, Enserro D, Xanthakis V, Aparicio H, Satizabal CL, et al. Association of Ideal Cardiovascular Health With Vascular Brain Injury and Incident Dementia. *Stroke* 2016

Peters F, Villeneuve S, Belleville S. Predicting progression to dementia in elderly subjects with mild cognitive impairment using both cognitive and neuroimaging predictors. *J. Alzheimers. Dis.* 2014; 38: 307–18.

Petersen RC, Aisen PS, Beckett LA, Donohue MC, Gamst AC, Harvey DJ, et al. Alzheimer's Disease Neuroimaging Initiative (ADNI): Clinical characterization. *Neurology* 2010; 74: 201–209.

Peterson D, Munger C, Crowley J, Corcoran C, Cruchaga C, Goate AM, et al. Variants in PPP3R1 and MAPT are associated with more rapid functional decline in Alzheimer's disease: The Cache County Dementia Progression Study *ALZHEIMERS Dement.* 2014; 10: 366–371.

Pimouguet C, Delva F, Le Goff M, Stern Y, Pasquier F, Berr C, et al. Survival and early recourse to care for dementia: A population based study *Alzheimer's Dement.* 2015; 11: 385–393.

Prince M, Wimo A, Guerchet M et al. World Alzheimer Report 2015: The Global Impact of Dementia | Alzheimer's Disease International *Alzheimer's Dis. Int.* London 2015

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira M a R, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 2007; 81: 559–575.

Querfurth HW, LaFerla FM. Alzheimer's disease. *N. Engl. J. Med.* 2010; 362: 329–344.

Ramanan VK, Risacher SL, Nho K, Kim S, Swaminathan S, Shen L, et al. APOE and BCHE as modulators of cerebral amyloid deposition: a florbetapir PET genome-wide association study. *Mol. Psychiatry* 2014; 19: 351–7.

Reeve A, Simcox E, Turnbull D. Ageing and Parkinson's disease: Why is advancing age the biggest risk factor? *Ageing Res. Rev.* 2014; 14: 19–30.

Roses AD. On the discovery of the genetic association of Apolipoprotein E genotypes and common late-onset Alzheimer disease. *J. Alzheimers. Dis.* 2006; 9: 361–6.

Roussotte FF, Gutman BA, Madsen SK, Colby JB, Thompson PM, Initiative for the ADN. Combined Effects of Alzheimer Risk Variants in the CLU and ApoE Genes on Ventricular Expansion Patterns in the Elderly *J. Neurosci.* 2014; 34: 6537–6545.

Rowe CC, Ellis K a., Rimajova M, Bourgeat P, Pike KE, Jones G, et al. Amyloid imaging results from the Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging *Neurobiol. Aging* 2010; 31: 1275–1283.

Sajjad A, Mirza SS, Portegies MLP, Bos MJ, Hofman A, Koudstaal PJ, et al. Subjective memory complaints and the risk of stroke. *Stroke.* 2015; 46: 170–5.

Scheltens P, Blennow K, Breteler MMB, de Strooper B, Frisoni GB, Salloway S, et al. Alzheimer's disease *Lancet* 2016; 388: 505–517.

Seshadri S, Fitzpatrick AL, Ikram MA, DeStefano AL, Gudnason V, Boada M, et al. Genome-wide analysis of genetic loci associated with Alzheimer disease. *JAMA* 2010; 303:

1832–1840.

Sevigny J, Chiao P, Bussière T, Weinreb PH, Williams L, Maier M, et al. The antibody aducanumab reduces A β plaques in Alzheimer's disease. *Nature* 2016; 537: 50–6.

Shaw LM, Vanderstichele H, Knapik-Czajka M, Clark CM, Aisen PS, Petersen RC, et al. Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging initiative subjects. *Ann. Neurol.* 2009; 65: 403–413.

Shi J, Lepore N, Gutman BA, Thompson PM, Baxter LC, Caselli RJ, et al. Genetic Influence of Apolipoprotein E4 Genotype on Hippocampal Morphometry : An N 5 725 Surface-Based Alzheimer's Disease Neuroimaging Initiative Study. *Hum. Brain Mapp.* 2014; 35: 3903–3918.

Silbert LC. Does statin use decrease the amount of Alzheimer disease pathology in the brain? *Neurology* 2007; 69: 8–12.

Singh N, Haldar S, Tripathi AK, McElwee MK, Horback K, Beserra A. Iron in neurodegenerative disorders of protein misfolding: a case of prion disorders and Parkinson's disease. *Antioxid. Redox Signal.* 2014; 21: 471–84.

Slegers K, Bettens K, De Roeck A, Van Cauwenberghe C, Cuyvers E, Verheijen J, et al. A 22-single nucleotide polymorphism Alzheimer's disease risk score correlates with family history, onset age, and cerebrospinal fluid A β 42. *Alzheimer's Dement.* 2015; 11: 1452–1460.

Smith R, Puschmann A, Schöll M, Ohlsson T, van Swieten J, Honer M, et al. ¹⁸F-AV-1451 tau PET imaging correlates strongly with tau neuropathology in *MAPT* mutation carriers *Brain* 2016: aww163.

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. *Alzheimer's Dement.* 2011; 7: 280–292.

Spulber G, Simmons A, Muehlboeck JS, Mecocci P, Vellas B, Tsolaki M, et al. An MRI-based index to measure the severity of Alzheimer's disease-like structural pattern in subjects with mild cognitive impairment. *J. Intern. Med.* 2013; 273: 396–409.

Steinberg M, Tschanz JT, Corcoran C, Steffens DC, Norton MC, Lyketsos CG, et al. The persistence of neuropsychiatric symptoms in dementia: The Cache County Study. *Int. J. Geriatr. Psychiatry* 2004; 19: 19–26.

Stephan BCM, Tzourio C, Auriacombe S, Amieva H, Dufouil C, Alperovitch A, et al. Usefulness of data from magnetic resonance imaging to improve prediction of dementia: population based cohort study. *BMJ* 2015; 350: h2863.

Strozyk D, Blennow K, White LR, Launer LJ. CSF A β 42 levels correlate with amyloid-neuropathology in a population-based autopsy study. *Neurology* 2003; 60: 652–656.

Swaen G, van Amelsvoort L. A weight of evidence approach to causal inference. *J. Clin. Epidemiol.* 2009; 62: 270–277.

- Takashima A. Amyloid-beta, tau, and dementia. *J. Alzheimer's Dis.* 2009; 17: 729–736.
- Tapiola T, Alafuzoff I, Herukka S-K, Parkkinen L, Hartikainen P, Soininen H, et al. Cerebrospinal fluid β -amyloid 42 and tau proteins as biomarkers of Alzheimer-type pathologic changes in the brain. *Arch. Neurol.* 2009; 66: 382–389.
- Teipel S, Heinsen H, Amaro E, Grinberg LT, Krause B, Grothe M. Cholinergic basal forebrain atrophy predicts amyloid burden in Alzheimer's disease. *Neurobiol. Aging* 2014; 35: 482–491.
- Therneau TM, Lumley T. Package 'survival'. 2016
- 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; 63110: 1–12.
- Toledo JB, Da X, Weiner MW, Wolk DA, Xie SX, Arnold SE, et al. CSF ApoE levels associate with cognitive decline and MRI changes. *Acta Neuropathol.* 2014; 127: 621–632.
- Trzepacz PT, Yu P, Sun J, Schuh K, Case M, Witte MM, et al. Comparison of neuroimaging modalities for the prediction of conversion from mild cognitive impairment to Alzheimer's dementia. *Neurobiol. Aging* 2014; 35: 143–51.
- Vemuri P, Wiste HJ, Weigand SD, Shaw LM, Trojanowski JQ, Weiner MW, et al. MRI and CSF biomarkers in normal, MCI, and AD subjects: Predicting future clinical change. *Neurology* 2009; 73: 294–301.
- Villemagne VL, Fodero-Tavoletti MT, Masters CL, Rowe CC. Tau imaging: Early progress and future directions. *Lancet Neurol.* 2015; 14: 114–124.
- Vivot A, Glymour MM, Tzourio C, Amouyel P, Chêne G, Dufouil C. Association of Alzheimer's related genotypes with cognitive decline in multiple domains: results from the Three-City Dijon study. *Mol. Psychiatry* 2015: 1–6.
- Vos SJB, Xiong C, Visser PJ, Jasielec MS, Hassenstab J, Grant EA, et al. Preclinical Alzheimer's disease and its outcome: A longitudinal cohort study. *Lancet Neurol.* 2013; 12: 957–965.
- Whitwell JL, Dickson DW, Murray ME, Weigand SD, Tosakulwong N, Senjem ML, et al. Neuroimaging correlates of pathologically defined subtypes of Alzheimer's disease: A case-control study. *Lancet Neurol.* 2012; 11: 868–877.
- Whitwell JL, Josephs KA, Murray ME, Kantarci K, Przybelski SA, Weigand SD, et al. MRI correlates of neurofibrillary tangle pathology at autopsy: A voxel-based morphometry study. *Neurology* 2008; 71: 743–749.
- Winblad B, Amouyel P, Andrieu S, Ballard C, Brayne C, Brodaty H, et al. Defeating Alzheimer's disease and other dementias: A priority for European science and society. *Lancet Neurol.* 2016; 15: 455–532.
- Xie L, Kang H, Xu Q, Chen MJ, Liao Y, Thiyagarajan M, et al. Sleep drives metabolite clearance from the adult brain. *Science* (80-.). 2013; 342: 373–377.

Young AL, Oxtoby NP, Daga P, Cash DM, Fox NC, Ourselin S, et al. A data-driven model of biomarker changes in sporadic Alzheimer's disease. *Brain* 2014a; 137: 2564–77.

Young AL, Oxtoby NP, Daga P, Cash DM, Fox NC, Ourselin S, et al. A data-driven model of biomarker changes in sporadic Alzheimer's disease. *Brain* 2014b

Van Der Zee J, Broeckhoven C Van. Dementia in 2013: Frontotemporal lobar degeneration—building on breakthroughs *Nat. Publ. Gr.* 2014: 2013–2014.

ADNI procedures manual ; 5: i–i. Available from: http://adni.loni.usc.edu/wp-content/uploads/2010/09/ADNI_GeneralProceduresManual.pdf

8. Appendix

8.1 MRI preprocessing

Image pre-processing consisted of a number of steps including unified segmentation, deformation with the Dartel algorithm. Images from ADNI and 3C were normalized into a template generated from the PPG to allow comparisons with that cohort. This step was followed by Jacobian modulation, smoothing (with an 8mm half-width Gaussian kernel) and spatial registration to MNI space. All pre-processed images passed through a visual quality control step before further analysis.

We first applied two-sample T-tests and identified regions with different distributions in grey matter volume between the 33 PPG and ADNI cohorts, adjusted by age, gender and total intracranial volume (TIV). To avoid cohort bias, we excluded these regions in the training and prediction process. We conducted within-subject standardization with mean zero and standard deviation one in all individuals from the three cohorts. Standardized images were adjusted for the effects of age and TIV using pathologically verified cognitively normal controls with a linear model.

8.2 Genetic variant processing

Among 20 AD related SNPs, 11 SNPs were successfully genotyped by the Illumina Human610 array in ADNI1 and 12 SNPs by the Omni 2.5M array in ADNI 2. SNPs unavailable in the genotyping panels used were replaced by the best available proxy SNPs whenever the r^2 was greater than 0.8 in 1000 Genomes Pilot 1 using SNAP, Broad (<http://www.broadinstitute.org/mpg/snap/index.php>). All SNPs showed a > 90% genotyping call rate, minor allele frequency >1% and Hardy-Weinberg Equilibrium P-value >1E-5. Measurements with missing values were imputed by the MACH algorithm.

8.3 Linear FA with expectation maximization (EM) algorithm

From equation (5.3), the cross-covariance between V and H can be identified:

$$\text{cov}(H, V) = E\left((H - \mathbf{0})(V - C)^T\right) = E(HH^T F^T) = F^T \quad (8.1)$$

Based on equation (8.1), the joint distribution between V and H is:

$$p(H, V) = p\left(\begin{bmatrix} H \\ V \end{bmatrix}\right) = N\left(\begin{bmatrix} H \\ V \end{bmatrix} \mid \begin{bmatrix} \mathbf{0} \\ c \end{bmatrix}, \begin{bmatrix} I & F^T \\ F & FF^T + \Psi \end{bmatrix}\right) \quad (8.2)$$

, where c is one column of C . The EM algorithm contains E step (expectation) and M (maximization) step. In the E step we calculate the conditional value of H using the joint distribution between V and H . Based on equation (8.2), the expected mean and variance of H can be expressed:

$$E(\mathbf{H}|\mathbf{V}) = \mathbf{F}^T(\mathbf{F}\mathbf{F}^T + \boldsymbol{\Psi})^{-1}(\mathbf{V} - \mathbf{C}) \quad (8.3)$$

$$\mathbf{V} = \mathbf{I} - \mathbf{F}^T(\mathbf{F}\mathbf{F}^T + \boldsymbol{\Psi})^{-1}\mathbf{F} \quad (8.4)$$

$$E(\mathbf{H}^T\mathbf{H}|\mathbf{V}) = \mathbf{V} + E(\mathbf{H}|\mathbf{V})^T E(\mathbf{H}|\mathbf{V}) \quad (8.5)$$

The values of \mathbf{F} and $\boldsymbol{\Psi}$ can be initialized with random values. To avoid local maximum, the values of the two parameters were initialized with the results from principal component analysis. The M step is to maximize the likelihood of data. The likelihood of \mathbf{V} is :

$$L(\mathbf{V}) = \prod_{n=1}^N p(\mathbf{V}_n|\mathbf{H})p(\mathbf{H}) \quad (8.6)$$

For computational continence, the maximization of equation (8.6) can be done by maximization of the log of the likelihood of \mathbf{V} .

$$\begin{aligned} l(\mathbf{V}) &= \log L(\mathbf{V}) \\ &= \text{con} - \frac{1}{2} \sum_{n=1}^N \log|\boldsymbol{\Psi}| - \frac{1}{2} \sum_{n=1}^N (\mathbf{V}_n - \mathbf{F}\mathbf{H} - \mathbf{C})^T \boldsymbol{\Psi}^{-1} (\mathbf{V}_n - \mathbf{F}\mathbf{H} - \mathbf{C}) \end{aligned} \quad (8.7)$$

To maximize equation (8.7), the values of \mathbf{F} and $\boldsymbol{\Psi}$ can be updated by setting the derivative of equation (8.7) to zero with respect to \mathbf{F} and $\boldsymbol{\Psi}$ respectively. Therefore, the updated \mathbf{F} and $\boldsymbol{\Psi}$ are:

$$\mathbf{F}^{new} = \sum_{n=1}^N ((\mathbf{V}_n - \mathbf{C})E(\mathbf{H}|\mathbf{V}_n)^T) \sum_{n=1}^N E(\mathbf{H}\mathbf{H}^T|\mathbf{V}_n)^{-1} \quad (8.8)$$

$$\boldsymbol{\Psi}^{new} = \frac{1}{N} \sum_{n=1}^N \{(\mathbf{V}_n - \mathbf{C})(\mathbf{V}_n - \mathbf{C})^T - \mathbf{F}^{new} E(\mathbf{H}|\mathbf{V}_n)(\mathbf{V}_n - \mathbf{C})^T\} \quad (8.9)$$

An iterative procedure was implemented until the log likelihood of \mathbf{V} is converged.

8.4 Item characteristic function

From equation (5.2), for the ratio of region i , $ratio_i$ can be expressed as:

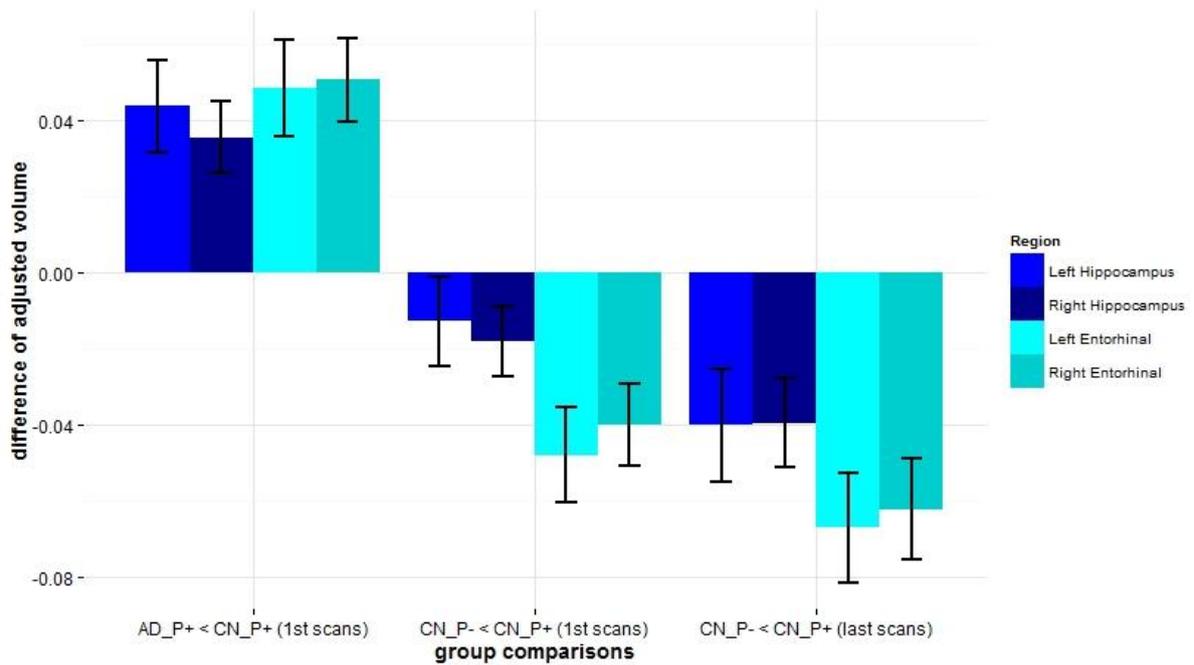
$$ratio_i = \frac{e^{V_i}}{1 + e^{V_i}} \quad (8.10)$$

The expectation of $ratio_i$, is:

$$E(\text{ratio}_i | H_j) = \int_{-\infty}^{\infty} \frac{e^{z\sigma_i + C_i + F_i H_j}}{1 + e^{z\sigma_i + C_i + F_i H_j}} \phi(z) dz, \quad (8.11)$$

where $\phi(z)$ is a standard normal density function and σ_i represents the standard deviation of ε_i in equation (5.3) and H_j indicates the j th row of the hidden factor matrix (the j th factor). The continuous distribution of z was approximated by a finite number of points from a discrete distribution.

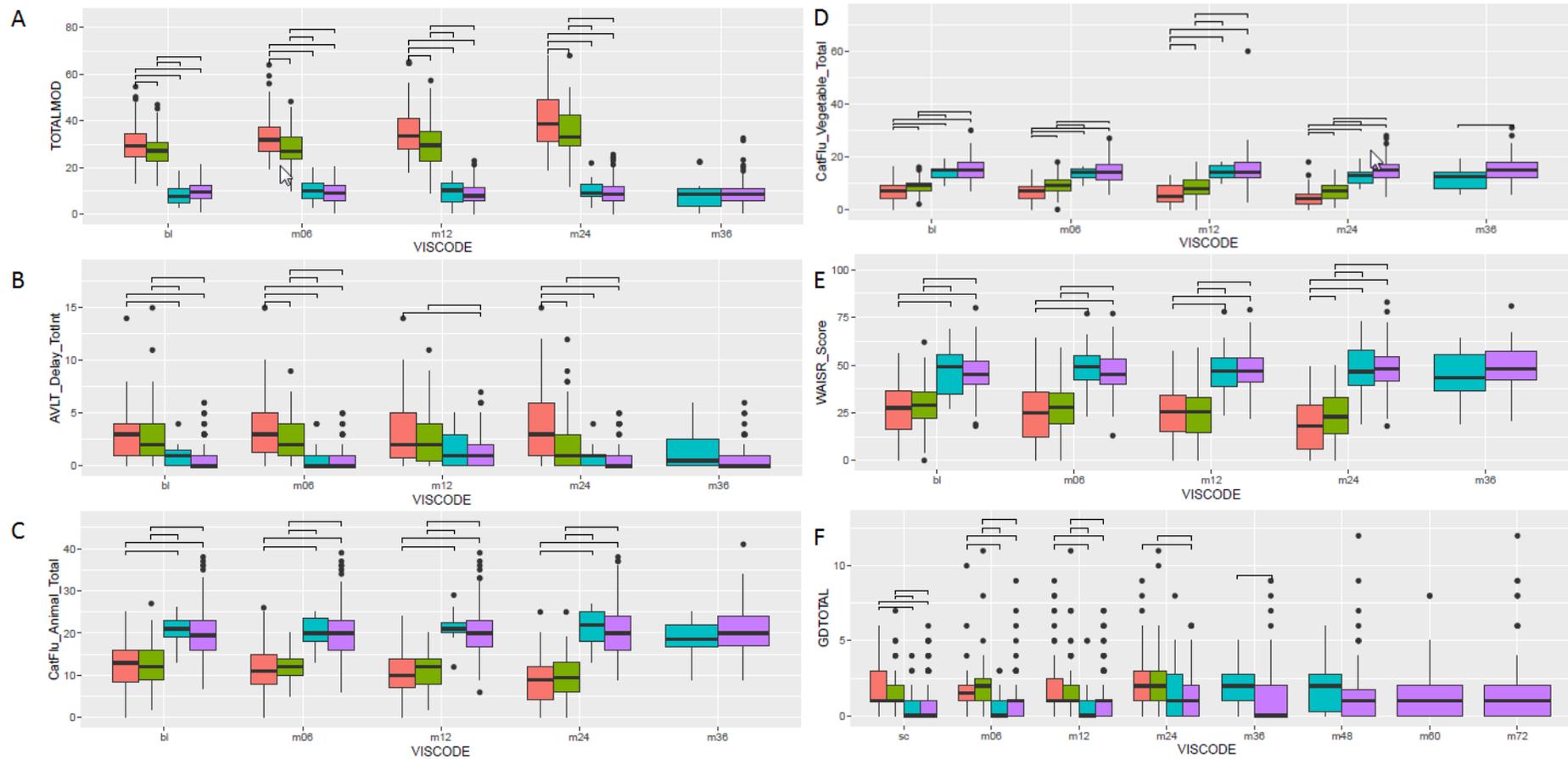
8.5 Supplementary Figure



Supplementary Figure 1 Volumetric difference between subgroups in hippocampal regions and entorhinal gyri

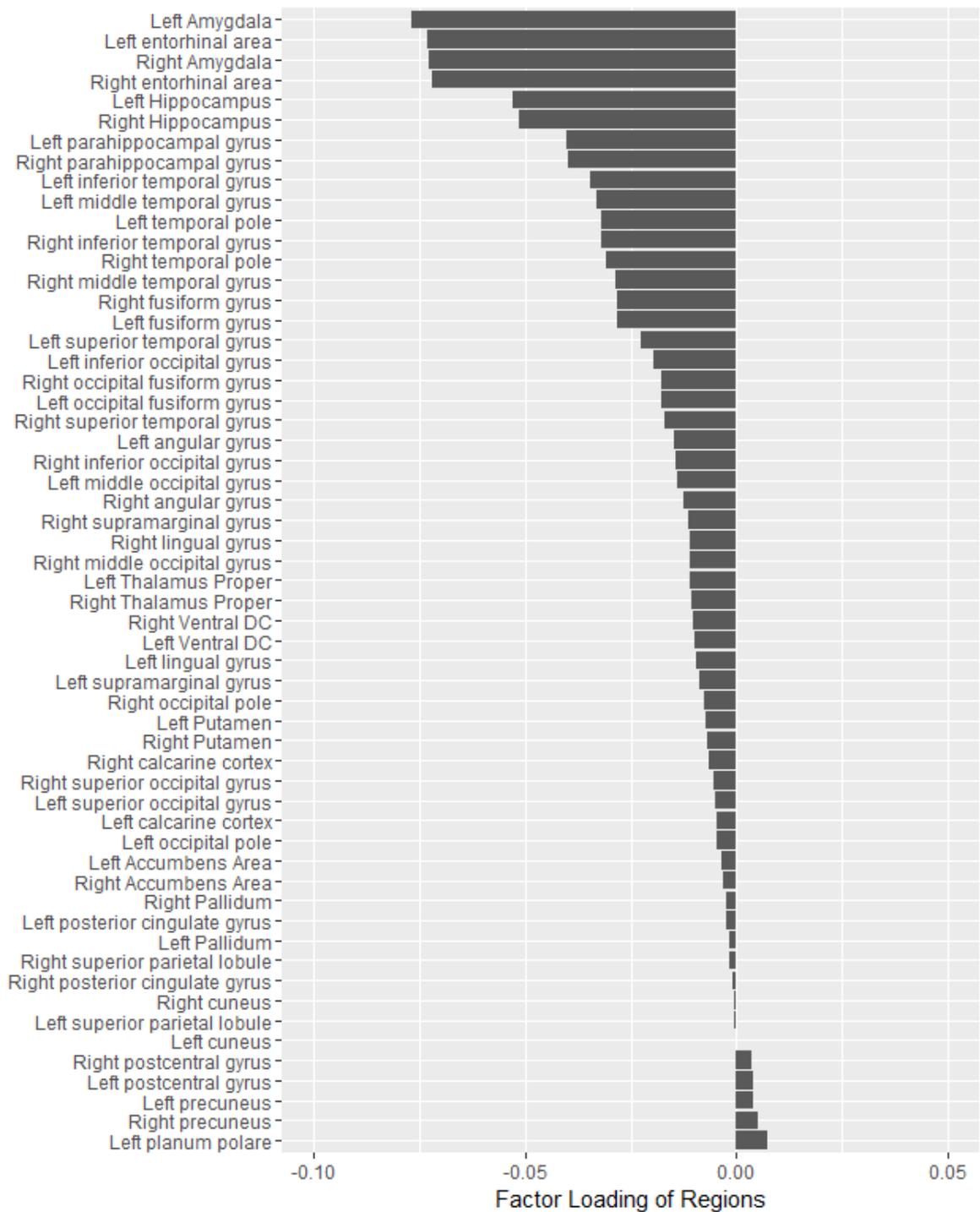
Average volume difference with standard error, after adjustment of age, gender and TIV, between predicted subgroups in ADNI. Four voxels are from left / right hippocampi and entorhinal cortices. CN_P- and CN_P+ were compared at baseline and with longitudinal scans.

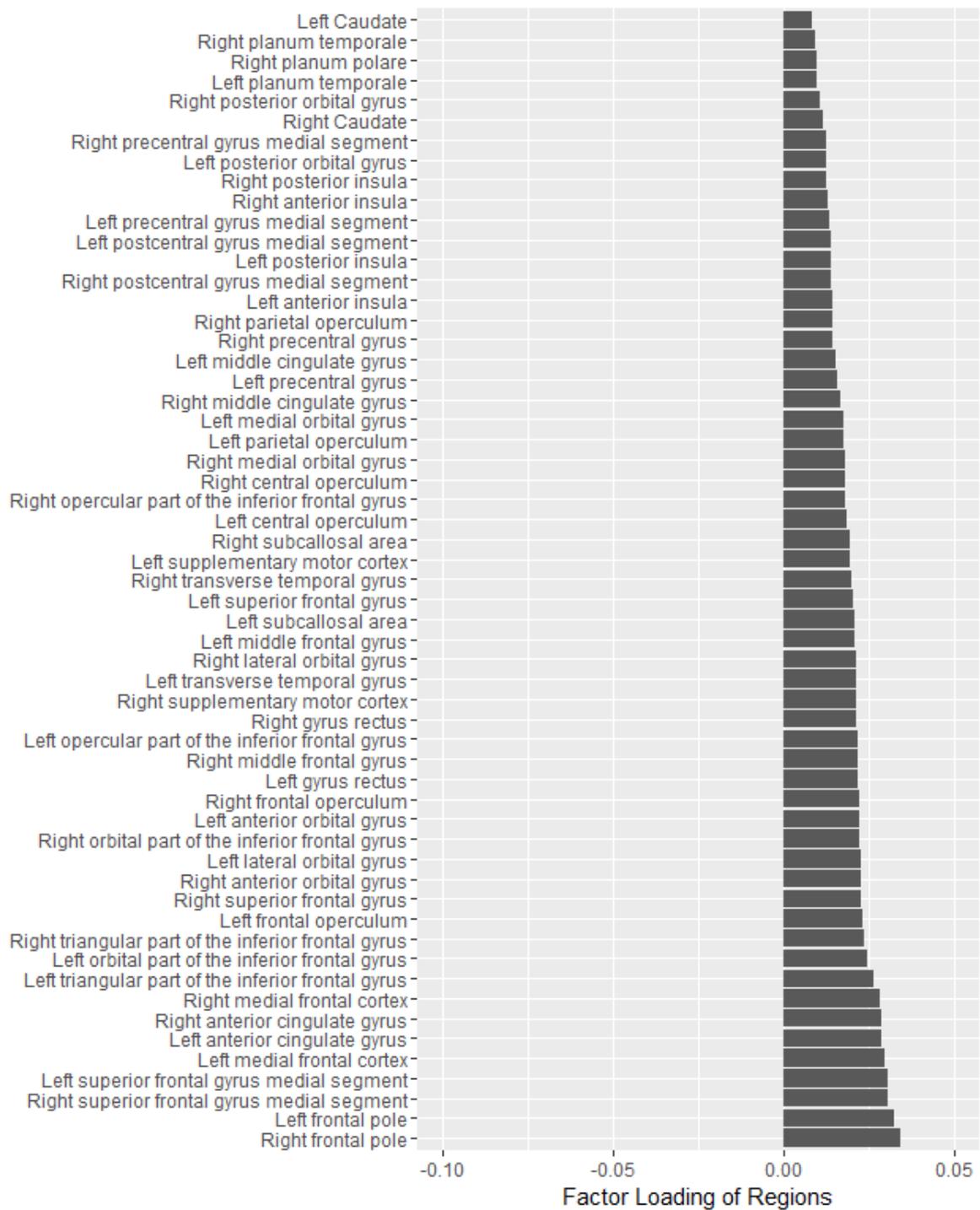
Coordinates: Left Hippocampus [-28.5,-31.5,-10.5], Left Entorhinal [-22.5,3,-27], Right Hippocampus [-32.3,-15,-16.9], Right Entorhinal [-30,3,-27]



Supplementary Figure 2 Clinical score distributions in follow-up study between ADNI predicted group

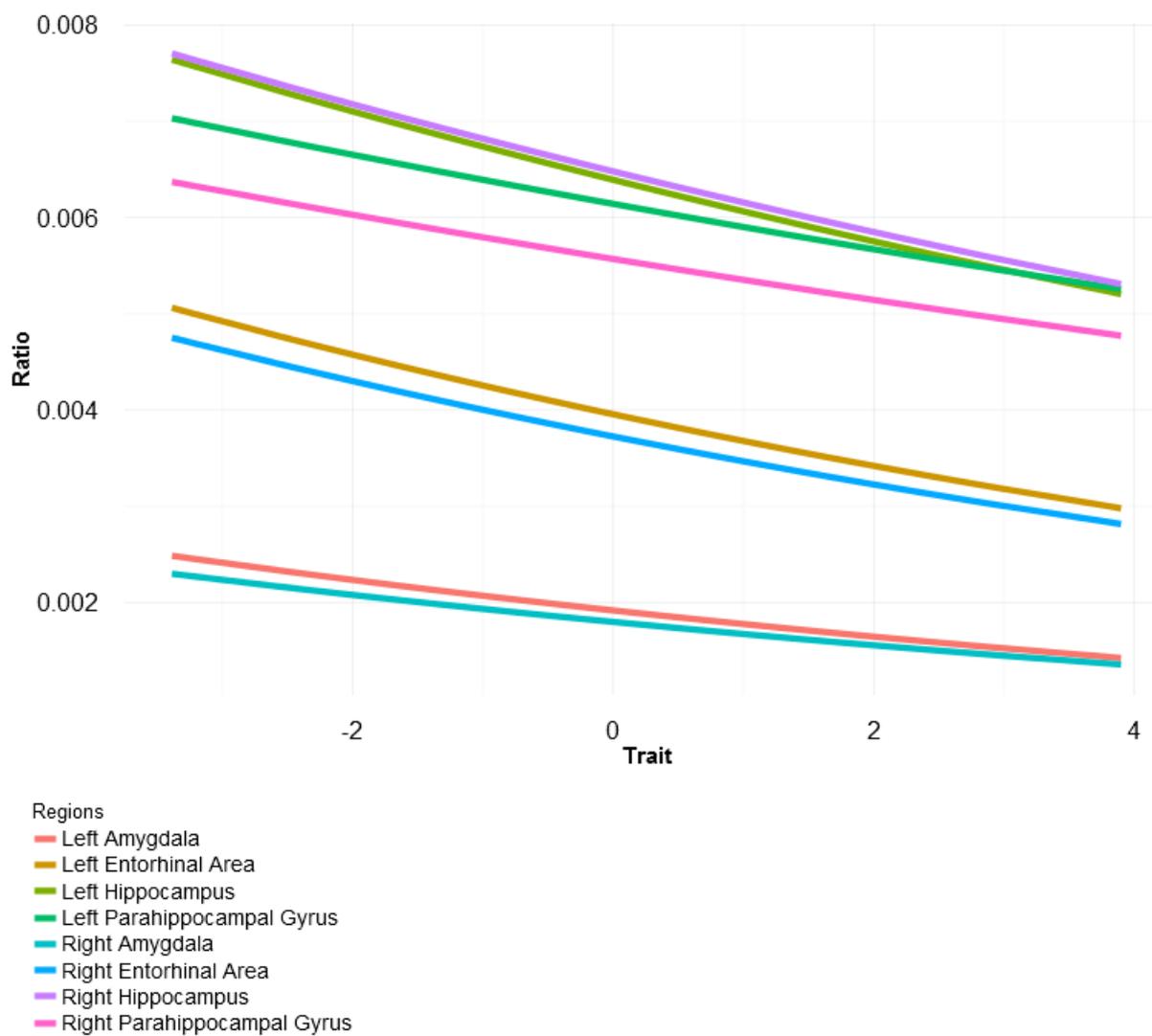
Clinical scores collected in ADNI study from baseline till 72 months later that are compared between each pair of groups. *TOTALMOD*: Alzheimer's Disease Assessment Scale-Cognitive (ADAS-Cog) Total Mod. *AVLT Delayed*: Auditory Verbal Learning Test. *Category Fluency Test*: Total number of animals and vegetables named. *WAIS-R*: Wechsler Adult Intelligence Scale-Revised (WAIS-R) Digit symbol substitution score. The statistical test was applied using Wilcoxon rank-sum test and distributions with significant differences were indicated by a horizontal bars above the relevant groups ($p < 0.05$).





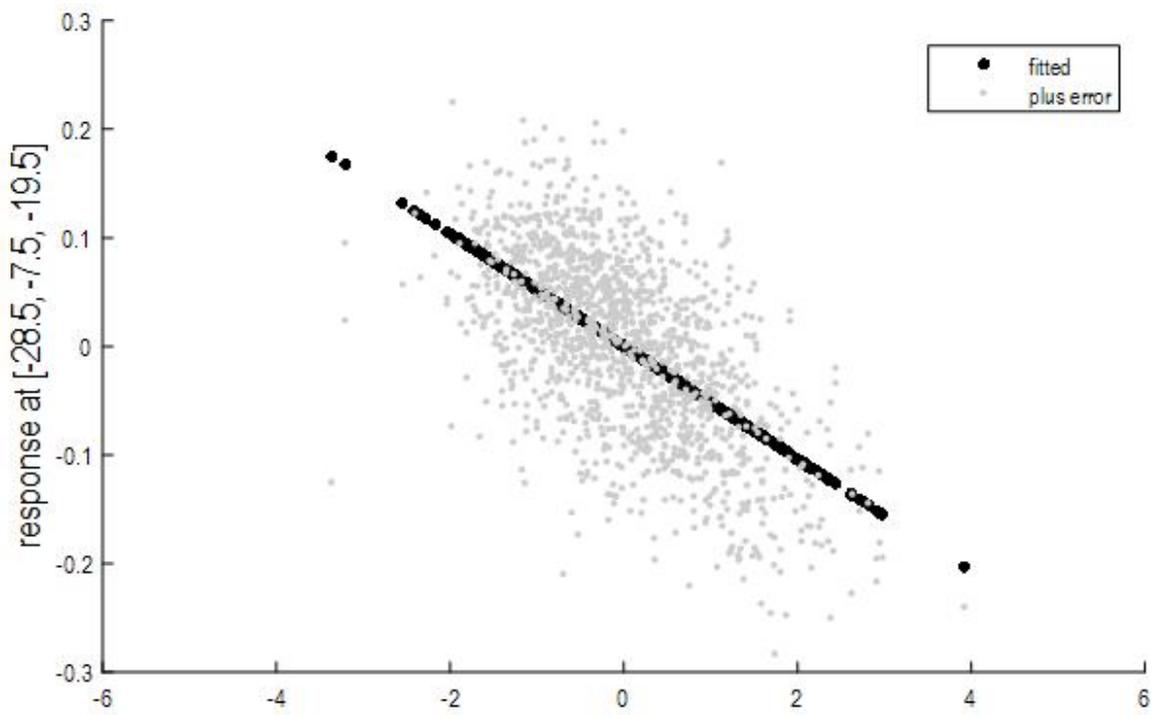
Supplementary Figure 3 Values of 114 regions in factor loading matrix

The factor loading matrix values of 114 regions indicate the covariance between the brain regions and the latent trait. The variable names are written on the vertical axis.

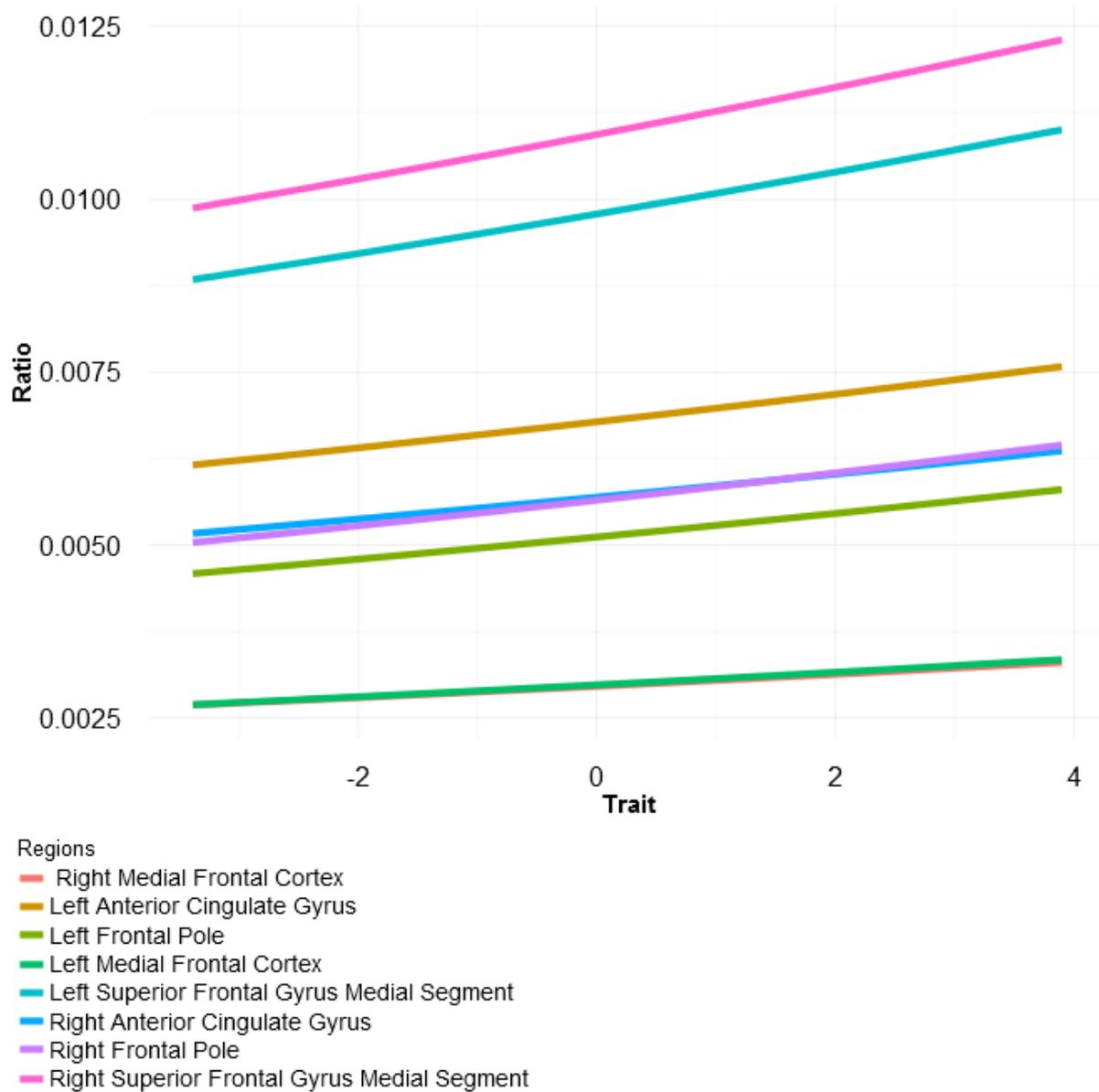


Supplementary Figure 4 Expected regional ratios as a function of the latent trait values

This figure shows the conditional regional ratios distribution for give latent trait. Regions of this figure are with lowest factor loadings (top 8 in Supplementary Figure 3).

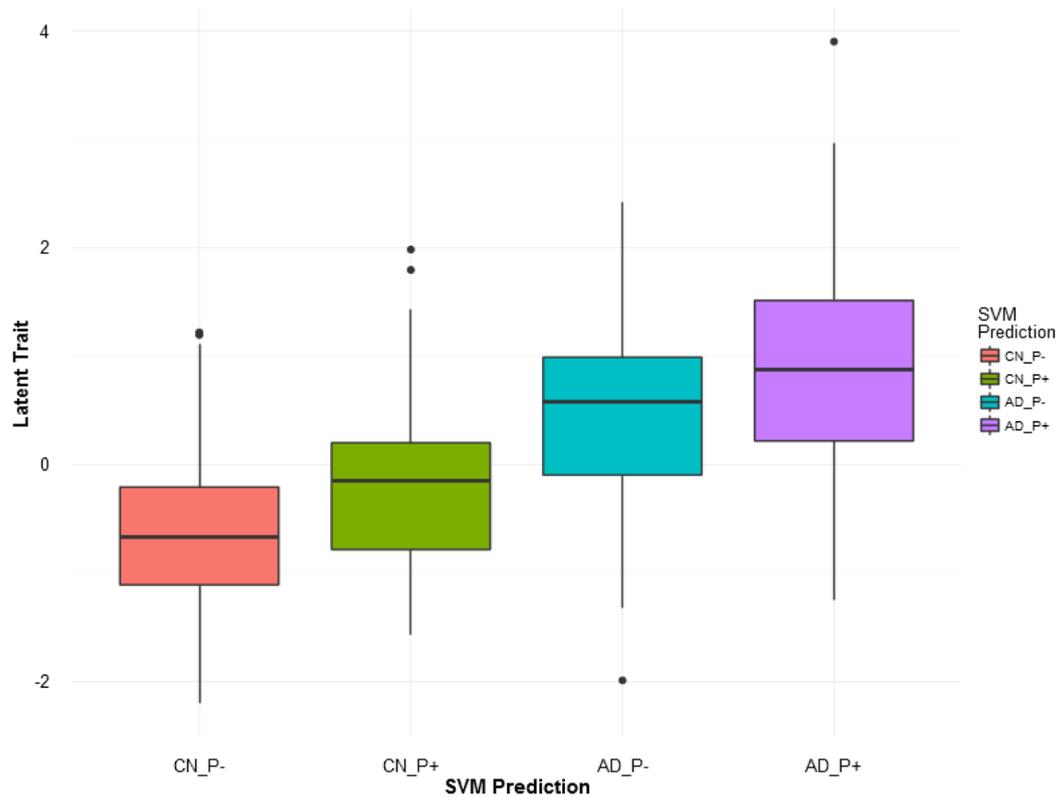


Supplementary Figure 5 Adjusted volume of one voxel of amygdale across the estimated latent trait



Supplementary Figure 6 Expected regional ratios as a function of the latent trait values

This figure shows the conditional regional ratios distribution for give latent trait. Regions of this figure are with lowest factor loadings (bottom 8 in Supplementary Figure 3).



Supplementary Figure 7 The estimated severity from FA compared with SVM predictions for ADNI subjects

This figure shows the distributions of estimated severity using FA algorithm among the predicted subgroups from SVM classifier (identification of ASR-AD project). Clinical diagnosed mismatched predicted groups, CN_P- vs CN_P+ and AD_P- vs AD_P+, have significantly different distributions ($p=0.0005$ and $p=0.004$, respectively by Wilcoxon rank sum test).

8.6 Supplementary Table

	Model 1 ^a ADNI CN to MCI or AD		Model 2 ^b ADNI CN to MCI		Model 3 ADNI MCI to AD	
	Hazard Ratio (95% CI)	P Value	Hazard Ratio (95% CI)	P Value	Hazard Ratio (95% CI)	P Value
Latent Trait	1.95(1.03-3.67)	0.04	2.15(1.09-4.22)	0.026	1.47(1.12-1.94)	0.006
Female	0.9(0.32-2.51)	0.833	0.78(0.26-2.3)	0.654	1.05(0.66-1.66)	0.838
Age	0.97(0.89-1.07)	0.579	0.98(0.89-1.08)	0.723	1.02(0.99-1.05)	0.315
APOE-ε4						
2	2.14(0.73-6.26)	0.164	2.51(0.87-7.25)	0.089	1.38(0.82-2.32)	0.228
1	7.67(1.18-49.73)	0.033	5.89(0.72-48.2)	0.099	1.12(0.55-2.3)	0.753
Education	0.89(0.74-1.07)	0.205	0.89(0.73-1.08)	0.226	0.96(0.89-1.03)	0.243
Tau	1.61(0.33-7.81)	0.553	1.88(0.4-8.84)	0.426	0.88(0.46-1.67)	0.696
pTau	1.64(0.35-7.8)	0.532	1.39(0.28-6.94)	0.687	1.61(0.81-3.22)	0.177
Aβ	0.6(0.16-2.27)	0.451	0.35(0.05-2.35)	0.282	0.66(0.31-1.41)	0.28

Supplementary Table 1 Effect of risk factors including CSF proteins on conversion from cognitive normal to MCI/AD and from MCI to AD in ADNI and 3C cohort

^aEvent-free survival is defined as the time of CN from baseline until the first clinical diagnosis of MCI, when available, or AD.

^bEvent-free survival is defined as the time of CN from baseline until the first clinical diagnosis of clinical AD.