Distinguishing early and late brain aging from the Alzheimer's disease spectrum: consistent morphological patterns across independent samples

Running title: Brain patterns distinguishing early and late aging from Alzheimer's disease

Authors: Nhat Trung Doan[a#], Andreas Engvig[a,b], Krystal Zaske[a], Karin Persson[c,d], Martina Jonette Lund[a], Tobias Kaufmann[a], Aldo Cordova-Palomera[a], Dag Alnæs[a], Torgeir Moberget[a], Anne Brækhus[c,d], Maria Lage Barca[c,d], Jan Egil Nordvik[e], Knut Engedal[c,d], Ingrid Agartz[a,f], Geir Selbæk[c,g], Ole A. Andreassen[a], Lars T. Westlye[a,h], for the Alzheimer's Disease Neuroimaging Initiative[*]

[a]*NORMENT, KG Jebsen Centre for Psychosis Research, Division of Mental Health and Addiction, Oslo University Hospital & Institute of Clinical Medicine, University of Oslo, Norway*,
[b]*Department of Medicine, Diakonhjemmet hospital, Oslo, Norway,*
[c]*Norwegian National Advisory Unit on Ageing and Health, Vestfold Hospital Trust, Tønsberg, Norway,*
[d]*Department of Geriatric Medicine, The Memory Clinic, Oslo University Hospital, Oslo, Norway,*
[e]*Sunnaas Rehabilitation Hospital HT, Nesodden, Norway,*
[f]*Department of Psychiatric Research, Diakonhjemmet Hospital, Oslo, Norway*
[g]*Centre for Old Age Psychiatric Research, Innlandet Hospital Trust, Ottestad, Norway,*
[h]*Department of Psychology, University of Oslo, Oslo, Norway*

Corresponding author: Nhat Trung Doan, Ph.D.
Email: n.t.doan@medisin.uio.no, Postal address: Oslo University Hospital, PoBox 4956 Nydalen, 0424 OSLO, Norway, Phone: +47 23 02 73 50, Fax: +47 23 02 73 33

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

Abbreviations: AD: Alzheimer's Disease, AUC: Area Under the receiver-operator characteristics Curve, FWHM: Full-width at Half Maximum, GLM: General Linear Model, HCO/HC: Healthy Control Older Adults, HCY: Healthy Control Young Adult, ICA: Independent Component Analysis, LICA: Linked Independent Component Analysis, MCI: Mild Cognitive Impairment, MMSE: Mini-Mental State Examination, MRI: Magnetic Resonance Imaging, MTL: Medial Temporal Lobe, SCI: Subjective Cognitive Impairment, GMD: Gray Matter Density, VBM: Voxel-based Morphometry.

Abstract

Alzheimer's disease (AD) is a debilitating age-related neurodegenerative disorder. Accurate identification of individuals at risk is complicated as AD shares cognitive and brain features with aging. We applied linked independent component analysis (LICA) on three complementary measures of gray matter structure: cortical thickness, area and gray matter density of 137 AD, 78 mild (MCI) and 38 subjective cognitive impairment patients, and 355 healthy adults aged 18-78 years to identify dissociable multivariate morphological patterns sensitive to age and diagnosis. Using the *lasso* classifier, we performed group classification and prediction of cognition and age at different age ranges to assess the sensitivity and diagnostic accuracy of the LICA patterns in relation to AD, as well as early and late healthy aging. Three components showed high sensitivity to the diagnosis and cognitive status of AD, with different relationships with age: one reflected an anterior-posterior gradient in thickness and gray matter density and was uniquely related to diagnosis, whereas the other two, reflecting widespread cortical thickness and medial temporal lobe volume, respectively, also correlated significantly with age. Repeating the LICA decomposition and between-subject analysis on ADNI data, including 186 AD, 395 MCI and 220 age-matched healthy controls, revealed largely consistent brain patterns and clinical associations across samples. Classification results showed that multivariate LICA-derived brain characteristics could be used to predict AD and age with high accuracy (area under ROC curve up to 0.93 for classification of AD from controls). Comparison between classifiers based on feature ranking and feature reduction suggests both common and unique feature sets implicated in AD and aging, and provides evidence of distinct age-related differences in early compared to late aging.

Keywords: Alzheimer's disease, Alzheimer's disease spectrum, early and late aging, linked independent component analysis, machine learning

1 Introduction

Sporadic Alzheimer's disease (AD) is a multifactorial neurodegenerative disorder strongly associated with increased age (Herrup, 2010), with more than 90% of all AD cases diagnosed after age 65 (Herrup, 2015). Normal aging and AD share cognitive and neuroanatomical characteristics in early stages (Fjell et al., 2013a; Fjell et al., 2014). Thus, discriminating between the two phenomena—as well as better deciphering their commonalities—is essential for disease-specific intervention and for stratifying groups in clinical trials.

Magnetic resonance imaging (MRI) may be used to study age-related and disease-specific patterns in the gray matter. Traditionally, atrophy of limbic structures, and particularly of the medial temporal lobe have been among the MRI features that distinguish best between patients with AD and healthy older adults (Fjell et al., 2010; Wang et al., 2015). Still, even for the established medial temporal lobe atrophy biomarker, accuracy is not ideal for individual AD diagnostics (Lowe et al., 2013), particularly in older age groups. Diagnostic accuracy of medial temporal lobe atrophy varies with the age of the patient, and may be complicated by differential effects of early and late brain aging processes (Fjell and Walhovd, 2010; Westlye et al., 2010).

The neurobiological mechanisms of MRI based cortical gray matter loss in both aging and across the dementia spectrum, including AD, are multifactorial—with e.g., functional, amyloidal, neurodegenerative and metabolic processes affecting brain structure differentially in different brain regions (Buckner et al., 2005). Since different biological processes that are distinguishable at the microscopic level may give rise to highly overlapping brain imaging features on the macroscopic level (e.g., cortical thickness), univariate methods, which only consider the observed level (e.g., cortical thickness in one region of interest or the volume of a subcortical structure)

and not the correlation structure across several MRI features (e.g., the association between cortical thickness in one region and hippocampal volume), may not be sensitive enough to properly differentiate aging and AD-specific processes. Instead, coordinated brain patterns caused by distinct underlying biological processes likely require multivariate approaches to disentangle (Doan et al., 2017; Doan et al., under review; Douaud et al., 2014; Francx et al., 2016; Groves et al., 2011).

Linked independent component analysis (LICA) is a promising multivariate technique for modeling co-variance across different brain indices or modalities (Groves et al., 2011). Unlike alternative supervised approaches such as partial least squares (Chen et al., 2009; Sui et al., 2012), which rely on the diagnosis label, LICA is fully data-driven and makes use of no demographic or diagnosis information. This technique was based on the conventional ICA technique, which assumes the signal to be a linear mixture of statistically independent spatial patterns that are non-Gaussian. During the optimization process, ICA searches for maximally non-Gaussian patterns by iteratively updating the subject loadings, or mixing parameters. LICA allows simultaneous ICA decompositions on different measures but constrains the subject loadings to be the same across measures (Groves et al., 2012). Given that each LICA component describes a spatial pattern of variation over and above variation associated with all other components, a particular attractive feature of LICA is that it allows identification of structured variance explaining only a small portion of the total variability, and which is therefore easily left unidentified when using conventional mass-univariate approaches (Doan et al., 2017; Francx et al., 2016).

Machine learning, or multivariate pattern analysis, offers a powerful option for building image-based predictive models useful for computer-aided diagnosis (Sabuncu et al., 2015; Westman et al., 2013; Westman et al., 2011). In AD, machine

learning has been used for diagnostic classification and clinical score prediction at an individual level (Cuingnet et al., 2011; Stonnington et al., 2010), outperforming radiological evaluation (Klöppel et al., 2012). A machine learning classifier, which can identify multivariate combinations of features that lead to maximal group classification or clinical score prediction accuracy, can be used to assess the clinical sensitivity of novel MRI features as a whole. Furthermore, the relative contribution, or ranking, of a feature used to build such a classifier with respect to all other features can also be evaluated to reveal further insights on the brain characteristics involved, and to compare the patterns involved in AD and healthy aging. The *lasso* (least absolute shrinkage and selection operator) algorithm, which has previously been successfully applied in clinical studies (Cai et al., 2014; Uddin et al., 2013; Wager et al., 2011), is efficient in identifying important features and yields shrinkage estimate of regression coefficients that potentially lower predictive errors compared to ordinary least squares (James et al., 2013).

With a primary interest in morphological feature extraction, we proposed to combine a multivariate data fusion approach with machine learning algorithms in an attempt to isolate and characterize MRI based brain morphometric features of the AD spectrum sharing commonalities with advancing age as well as disease-specific patterns. Rather than building a classification model of AD with maximal accuracy, we used machine learning as a means to further evaluate the extracted patterns for their differential sensitivity to disease and aging. Specifically, we performed LICA on three sensitive and complementary MRI-derived gray matter morphological measures (modulated gray matter density maps (GMD) obtained using voxel-based morphometry (VBM), cortical surface area and thickness) in a combined sample comprising 355 healthy subjects covering large portions of the adult human lifespan

(18 to 78 yrs) and 253 memory clinic patients with increasing degree of cognitive impairment (subjective cognitive impairment (SCI), mild cognitive impairment (MCI)) and AD dementia. In order to assess the generalizability of the multivariate patterns and their diagnostic sensitivity across cohorts, we attempted to replicate our findings in an independent dataset of well age-matched groups, comprising 186 patients with AD, 395 patients with MCI and 220 healthy controls (HC) from the Alzheimer's Disease Neuroimaging Initiative (ADNI). We included both cortical surface area and thickness in the model as they are genetically independent (Panizzon et al., 2009; Winkler et al., 2010) and provide complementary information on cortical morphology (Hogstrom et al., 2013; Hutton et al., 2009). GMD maps provide an indirect measure of gray matter volume enabling assessment of subcortical structures, such as the hippocampus. Furthermore, surface area and thickness contributes to only a proportion of the variance in GMD, as shown in schizophrenia studies (Palaniyappan and Liddle, 2012). Thus jointly analyzing the three measures would likely increase the sensitivity to clinical variance, and also provide more nuanced information on regional brain patterns.

In line with recent studies using LICA, we expected one or more widespread components capturing global cortical features with strong relationships with age (Douaud et al., 2014; Groves et al., 2012). By including an adult lifespan sample as well as a broad range of patients with AD and its clinical precursors, we investigated the assumption that AD is both associated with aging-related, as well as disease-specific brain patterns. We hypothesized that LICA would allow for identification of both common and differential features of AD and aging. Further, we hypothesized that feature sets informative for AD classification would be more similar to the feature sets involved in late compared to early aging. To test the clinical sensitivity of

the LICA multivariate patterns at an individual level, and to assess their relative involvement in AD and healthy aging, we used machine learning to perform pairwise clinical group classification, and prediction of cognition and age at different age ranges. Finally, we compared the feature importance in the resulting clinical group, cognition and age classifiers to assess the brain morphological overlap between these different phenotypes.

2 Materials and Methods

2.1 Participant recruitment and screening

For our discovery sample, cross-sectional patient data were obtained from the "Norwegian registry for persons being evaluated for cognitive symptoms in specialized care (NorCog)". NorCog is a national patient registry comprising consecutively enrolled patients referred to one of 27 participating memory outpatient clinics for workup of suspected cognitive impairment or dementia. The patients in the present study were recruited from one of the centers, the memory clinic at Oslo University Hospital between 2010 and 2014. Included patients were assessed in accordance with an extensive standardized clinical examination protocol (Braekhus et al., 2011), and referred to the same brain MRI as healthy controls. Two experienced memory clinic physicians diagnosed the patients according to research criteria in consensus (K.E./A.B., or M.L.B./K.P.). Only patients fulfilling *International Classification of Diseases, 10th Revision* criteria for AD (N=137; World Health Organization., 1993), and the Winblad criteria for MCI (N=78; Winblad et al., 2004), as well as patients referred with a subjective cognitive complaint that did not fulfill MCI or dementia criteria, termed SCI (N=38; Garcia-Ptacek et al., 2014) were included. Degree of cognitive impairment was quantified using the results of the MMSE from the clinical assessment. Additional descriptive information is summarized in Table 1.

Healthy controls were included retrospectively from two concurrent projects (STROKEMRI and TOP) using the same MRI scanner and pulse sequences as the patients. For STROKEMRI (Dørum et al., 2016; Dørum et al., 2017), healthy controls (18-78 yrs, *N*=52) were recruited through a newspaper ad and social media. Exclusion criteria included estimated IQ < 70, previous history of alcohol-and substance abuse, history of neurologic or psychiatric disease, participants presently on any medication significantly affecting the nervous system and contraindications for MRI. All participants were self-sufficient and living independently, and reported no reason to suspect marked cognitive decline or undiagnosed dementia. For TOP, healthy controls (18-46 yrs, *N*=303) were invited after a stratified random selection drawn from the Norwegian National Population Registry. All underwent initial interview where demographic and clinical information was obtained. Exclusion criteria included a history of head trauma with loss of consciousness of more than 10 minutes duration, moderate to severe psychiatric or somatic disease, first-degree relatives with mental illnesses (schizophrenia, bipolar disorder, and major depression disorder), excessive substance abuse during the last 12 months, or not being able to perform an MRI scan. Blood samples were taken for standard hospital hematological screening to rule out on-going illnesses and a urine sample was collected to screen for substance abuse. No MMSE examination was performed for the healthy controls, but all subjects underwent neuropsychological screening. For post-hoc and classification analyses, and also for data visualization the healthy controls were divided into two groups: healthy young adults (18 yrs < HCY < 45 yrs), and healthy middle-aged and older adults (HCO >= 45 yrs, Table 1) such that the HCO group was both of reasonable size and as age-matched as possible to the clinical groups (MCI, SCI and AD).

For the replication sample, we included age-matched groups from the ADNI1 cohort (AD: *N*=186, mean age = 75.2 ± 7.5 yrs, range 55-91 yrs, N_{female}=89; MCI: *N*=395, mean age

= 74.7 ± 7.4 yrs, range 54-89 yrs, N_{female}=140; HC: N=220, mean age = 75.9 ± 5.1 yrs, range 60-90 yrs, N_{female}=108). These data were obtained from the 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 MRI, positron emission tomography, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD. More details about the ADNI sample can be seen in (Jack et al., 2008; Petersen et al., 2010; Weiner et al., 2010).

2.2 *MRI acquisition*

A 3 Tesla GE Signa HDxT scanner at Oslo University Hospital was used to collect MR data using two different head coils (8-channel head coil (8HRBRAIN) and the Head/Neck/Spine (HNS) coil, *N* per group is given in Table 1). A T1-weighted 3D Fast Spoiled Gradient Echo (FSPGR) sequence was used with the following parameters: repetition time (TR) = 7.8 ms, echo time (TE) = 2.956 ms, inversion time (TI) = 450 ms, flip angle 12°, matrix = 256 x 256 mm, in-plane resolution=1x1mm, slice thickness=1.2mm; acquisition time=7min 8s, 166 sagittal slices. Details regarding MRI acquisition of the ADNI sample can be seen in (Jack et al., 2008).

2.3 *Image preprocessing*

T1-weighted scans were processed using FreeSurfer 5.3 (http://surfer.nmr.mgh.harvard.edu) to estimate vertex-wise cortical thickness and surface arealization (Dale et al., 1999). All datasets included in this study passed a rigorous quality control procedure, which included visual assessment of the segmentations, minor manual intervention to correct for segmentation errors wherever deemed applicable, and exclusion of datasets with significant low quality due to e.g., motion artifacts. Surface maps were resampled to a common coordinate system (fsaverage5, 10242 vertices) using a non-rigid high-dimensional spherical

averaging method to align cortical folding patterns (Fischl and Dale, 2000). Cortical thickness and surface area maps were smoothed using a Gaussian kernel with a commonly used full width of half maximum (FWHM) of 15 mm and 10 mm, respectively. Total hippocampus volume was also estimated based on FreeSurfer subcortical segmentations for post-hoc correlation analysis with the components implicating this structure.

GMD maps were derived using FSL-VBM (Douaud et al., 2007), an optimised VBM protocol (Good et al., 2002) carried out with FSL tools (Smith et al., 2004). First, structural images were brain-extracted and gray matter-segmented before being registered to the MNI152 standard space using non-linear registration (Andersson et al., 2007). The resulting images were averaged and flipped along the x-axis to create a left-right symmetric, study-specific grey matter template. Second, all native grey matter images were non-linearly registered to this study-specific template and "modulated" to correct for local expansion (or contraction) due to the non-linear component of the spatial transformation. The modulated gray matter maps were smoothed with a sigma of 4 mm (FWHM=9.4 mm). Note that although GMD-maps are interpreted as measure of cerebral GM, the FSL-VBM GM-segmentation did include white matter (WM)-voxels as revealed by manual inspection and overlap with template atlases, thought to result from the probability based nature of the segmentation scheme itself and age-related decreases and blurring of GM and WM contrast.

For a comparison with the multivariate features, we summarized the cortical thickness, surface area using FreeSurfer's Desikan- Killiany atlas (Desikan et al., 2009) and the GMD maps using the AAL atlas (Tzourio-Mazoyer et al., 2002) (mean value for thickness, GMD and sum for surface area within each ROI and whole brain), resulting in a total set of 262 univariate features, hereafter referred to as FS-VBM feature set. In addition, we performed a principle component analysis (PCA) on the FS-VBM set after normalizing all features (mean-centering and scaling to have a standard deviation of 1). We then kept the

PCA components (n=136) that explained in total 95% of the variance. The FS-VBM and PCA feature sets were used in multivariate group classification and age prediction (see below and in Supplemental information (SI)).

2.4 Linked independent component analysis

We performed a data-driven decomposition of the imaging features into independent components using FMRIB's LICA (FLICA, http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FLICA), which models the inter-subject variability across measures (Groves et al., 2011; Groves et al., 2012). A LICA component, characterized by its spatial maps and the subject weights that are shared across measures, may involve multiple or only one measure. A model order of 50 was chosen based on previous studies (Francx et al., 2016; Groves et al., 2012), resulting in a biologically meaningful yet manageable set of patterns. We assessed the effect of model order on the multivariate group classification and age prediction described below. Furthermore, at different model orders, we performed exploratory hierarchical clustering of the subjects and evaluated the resulting clusters using the cophenetic correlation coefficient (Farris, 1969). The results suggest that a model order of 50 was a suitable choice for the data (more details in SI). After visual inspection of the spatial maps, we excluded three components that showed strong head coil effect from further multivariate analyses. Additionally, we performed a comparison between ICA decomposition using either thickness, surface area or GMD maps, and LICA using all three measures simultaneously and presented the results in SI. Briefly, empirical results with the LICA components showing increased sensitivity to AD and superior classification performance in most cases (Fig. S10) provided supporting evidence for the benefits of combining these complementary measures using LICA.

2.5 Statistical analyses

2.5.1 Univariate analyses

Diagnostic associations with each of the component's subject loadings were tested using general linear models (GLM), and partial correlations with MMSE were tested, covarying for age (using second order orthogonal polynomials to account for linear and quadratic effects), sex and head coil (both as factorial variables). Effect size of group pairwise comparisons in subject loadings was standardized using the Cohen's *d* as follows: $Cohen's\ d = \frac{2*t}{\sqrt{df}}$, where *t* was the *t* statistics and *df* the degree of freedom of the residuals. We also tested for main effects of age and head coil in the same GLM. For the components showing significant age effects, we performed local polynomial fitting (LOESS) (Weisberg, 2005) of age against the subject loadings and computed R^2 as a measure of goodness of fit. The components capturing strong head coil effects were visually inspected; those reflecting non-anatomical spatial maps were considered noise and excluded from subsequent analyses.

2.5.2 Multivariate classification and predictions

To evaluate clinical sensitivity of the derived multivariate patterns at an individual level, we submitted the LICA features (*N*=47 after QC) to pairwise classification among the AD, MCI, SCI and HCO groups using the *lasso* classifier as implemented in the *glmnet* R package (Friedman et al., 2010). To account for effects of normal aging while avoiding removing disease-related effects associated with advancing age (Doan et al., 2017; Dukart et al., 2011; Koutsouleris et al., 2015), prior to classification, we estimated the age effects using the combined sets of HCY, HCO and SCI subjects by means of GLM with age (second order orthogonal polynomial) as the only independent variable. The resulting GLM model was used to compute the residuals of each feature on all datasets, including MCI and AD. We merged the SCI with the control groups to increase the sample size, especially at the elderly range, since classification between SCI and HCO showed chance level accuracy, indicating that

there were no notable neuroimaging phenotypic differences between these groups. We then normalized each feature such that it had zero mean and a standard deviation of 1.

In an attempt to disentangle the differential involvement of the LICA features in presumably healthy aging and dementia, respectively, we applied the *lasso* regression (Friedman et al., 2010) to predict age at different age ranges and compared the feature ranking obtained from the age prediction to the ranking obtained from pairwise group classifications. To be able to address the relation between dementia and aging in terms of the brain structural underpinnings, we only used subjects from the HCO, HCY and SCI groups for age prediction. We split these subjects into two groups at beginning of middle age (45 years, according to the National Library of Medicine's definition of middle age): one group with an age range of [18;45 yrs]), referred to as the early aging group, and one with an age range of [45;90 yrs], late aging. We performed age prediction on both early, late and full age range groups. We also used the *lasso* regression to predict MMSE within each of the clinical groups (AD, MCI, SCI). Whereas the raw features were used in age prediction, to account for possible confounding effects of normal aging in MMSE prediction, we used the same residualized features as those used for group classification.

To compare the involvement of the LICA features across classifiers, for each group or age classifier, we ran the same classification (or prediction) after incrementally removing a number of features (n=1,2,3,4,5 etc.…). The features excluded in each iteration were determined based on feature importance information derived from a reference classifier. For instance, when running AD *versus* HCO classification referring to the Age_18_45 (early) age classifier, we (1) ordered the LICA feature using the feature importance (standardized regression coefficient values) returned by the Age_18_45 classifier, (2) incrementally removed features starting from the most important features, ran AD *versus* HCO classification and computed the performance, (3) repeated the above procedure for each of

the classifiers (including AD *versus* HCO) as a reference classifier. If the classifiers were similar in terms of the features involved, then the performance would have a similar decay profile, and vice versa different performance profiles would be observed if they were different.

To assess the predictive power of each of the components sensitive to diagnosis in terms of group classification and age prediction compared to using all LICA features, we repeated the same classification and age prediction as described above. However, instead of using all LICA features, we used each of those components as the only feature and the linear regression classifier.

We used *k*-fold cross-validation to estimate the classification performance. For each classification, all datasets involved were split into *k* partitions of equal size. One partition was left out for validation (the testing set). The classifier was then built using the (*k-1*) remaining partitions, on which another *k*-fold cross-validation was applied to estimate the regularization parameter λ. As the group size was highly imbalanced in most cases, to alleviate classification bias toward higher accuracy on the majority class, we balanced class size of the training set (the testing set remained untouched), based on a resampling technique as implemented in the *ROSE* R package (Lunardon et al., 2014). The trained classifier was then applied on the left out partition. This process is repeated for each of the partitions. We chose the commonly used value of 10 for *k* (James et al., 2013). The entire process was repeated *N* times (*N*=100, chosen as an arbitrarily large number) and the average performance (balanced accuracy (Brodersen et al., 2010), accuracy, specificity, sensitivity, and area under the receiver-operator characteristics curve (AUC)) was computed. We inferred the relative feature importance (ranking) in a classification using the magnitude of the standardized regression coefficients averaged across folds and repetitions. The same *k*-fold cross-validation setup was used for group classification and age/MMSE prediction. All models

built for group classifications, age and MMSE predictions are coherently referred to as classifiers.

Lastly, for a comparison between multivariate LICA features and univariate FS-VBM, PCA features, we applied the same group pairwise classification and age prediction as described using either the FS-VBM or PCA feature sets and compared the performance obtained with LICA features to the resulting performance. The results are presented in SI.

LICA decomposition was performed in Matlab (version R2014a). All statistical and machine learning analyses were performed in R (http://cran.r-project.org, version 3.2.1)). The *glmnet* (Friedman et al., 2010) and *caret* (Kuhn, 2008) R packages were used for classification and prediction, and the *ggplot2* package (Wickham, 2009) for visualization. We corrected for multiple statistical comparisons using permutation testing (details are presented in SI) and used a significance threshold of 0.05. The *p*-values reported throughout the manuscript are corrected based on permutation testing, unless stated otherwise.

2.6 Replication analyses using ADNI

The T1-weighted images of the replication sample were processed using the same preprocessing pipeline, including LICA decomposition (model order of 50), as described above. Subsequently, we assessed the similarities between the resulting LICA patterns and those obtained from the discovery sample by correlating the corresponding spatial maps. We repeated the GLM analyses to study the main effect of diagnosis, accounting for age and sex. Further, we also repeated the classification analyses between AD, MCI and HC using the nested *k*-fold cross-validation framework as described above.

2.7 *Ethics*

The Regional Committee for Medical Research Ethics in South-Eastern Norway approved the study. All participants gave written informed consent. Patients were only enrolled if determined to have capacity for consent by the evaluating physician.

3 Results

Results on the main effects of diagnosis and age on the components' subject loadings are presented in Fig. 1 and Fig. S1 for the discovery and replication samples, respectively. Results for pairwise group comparison for the two samples in terms of Cohen's *d* are detailed in Fig. S2. Fifty independent components (IC) were derived for both samples. Unless otherwise stated, the components' indices presented and discussed refer to those obtained from the discovery sample.

3.1 Three spatial components relate to clinical diagnosis, with consistent patterns across independent samples

IC0, IC5, and IC9 (spatial maps and subject loading distribution presented in Fig. 2A-B), explaining 24.4%, 1.7% and 1.39% of the data variance (Fig. S3), respectively, showed strong effects of diagnosis. $IC3_{rep}$, $IC4_{rep}$, $IC8_{rep}$ (*rep* = replication sample; spatial maps and subject loadings presented in Fig. 2C-D) also showed significant diagnosis effects (Fig. S1) and resembled IC0, IC5 and IC9 (Fig. 2A-B), respectively, in the discovery sample ($IC3_{rep}$ *versus* IC0: $r_{thickness}$=0.66, r_{GMD}=0.69; $IC4_{rep}$ *versus* IC5: $r_{thickness}$=0.95, r_{GMD}=0.74; $IC8_{rep}$ *versus* IC9: $r_{thickness}$=0.57, r_{GMD}=0.3, Fig. 2A-C). We found a consistent graded pattern of AD < MCI < HCO, SCI or HC across cohorts, as visually indicated by the boxplots in Fig. 2.

We note that IC4 and IC37 also showed significant main group effects (Fig. 1) in the discovery sample. However, since IC4 showed a strong effect of head coil (F=36.9, Fig. S3) and both IC4 and IC37 did not show significant group pairwise differences (Fig. S2), we chose to withhold these components from further univariate analysis.

IC0 reflected global thickness variation co-occurring with GMD variation in opposite weightings (positive at subcortical structures, the lingual- occipital cortex, and negative at frontal white matter and temporal cortex; AD < MCI, SCI, HCO). IC5 showed a pattern of lower medial temporal lobe (MTL) thickness and hippocampal reductions in AD compared to

MCI and SCI (d=-0.27, -0.32, respectively). This component showed an association with hippocampus volume with a significant group by hippocampus volume interaction, showing the steepest slope within AD (r_{AD}=0.39, p=1.809x10^{-06}; r_{MCI}=0.05, p=0.67; r_{SCI}=0.21, p=0.2; r_{HCO}=0.04, p=0.83; r_{HCY}=0.12, p=0.03, accounting for age, sex and estimated intracranial volume).

IC9 revealed an anterior-to-posterior graded pattern reflecting relatively decreasing thickness and GMD along the posterior-anterior axis, representing structural variation over and above all other components (AD < MCI, SCI, HCO, Cohen's d = -0.64, -0.56, -0.61, respectively). The observed lower subject loadings in AD compared to controls reflect a shift of the thickness and GMD distribution towards a stronger anterior compared to posterior weighting. The positively weighted regions largely comprised the lateral posterior aspects of the temporal lobe, precuneus, and posterior cingulate cortex, indicating greater cortical atrophy in these regions in AD.

3.2 The diagnosis-related components show variable relationships with age

Fig. 3 plots subject loadings as a function of age for IC0, IC5, IC9 as well as for the corresponding IC3$_{rep}$, IC4$_{rep}$, IC8$_{rep}$ of the replication sample. Age scatter plots for other components in the discovery sample (IC1,3,7 and 14), which showed significant main effects of age (defined as $p < 0.05$; $R^2 \geq 0.1$), are presented in Fig. S4. In the discovery sample, among the three diagnosis-related components, IC0 showed a strong monotonic decrease with age (R^2=0.83 across all groups and 0.68 across HCO, HCY and SCI), reflecting cortical thinning and bidirectional (decrease-increase) GMD alterations as a function of age. IC5 on the other hand showed quadratic effect of age (F=35.8, p<0.001) with a modest fit of R^2=0.13 across all groups. Notably, the fitted curve across HCO, HCY and SCI (Fig. 3A, blue) depicted an increase in subject loadings from early until middle age. The curve then saturated during middle to late age. When additionally including MCI and AD in age fit, the curve

(red) decreased during middle to late age, suggesting a specific involvement of AD on IC5 in late aging.

IC9 was not significantly related to age (F = 4.8, p=0.35, R^2=0.01). Interestingly, whereas there was no significant correlation between IC9 and age within MCI, SCI or HCO (t=-1.8, 0.9, 1.7, Cohen's d=-0.42, 0.3, 0.64, p>0.05, respectively), there was a trend of positive association within the AD group (t=3.6, d=0.63, p=4.5x10^{-4}, uncorrected) indicating that the younger AD patients showed larger difference in subject loadings with respect to the other groups than the older AD patients (Fig. 3B).

In the replication sample, whereas IC3$_{rep}$ was strongly correlated with age (F=262, p=1x10^{-5}), IC4$_{rep}$ and IC8$_{rep}$ only showed moderate or weaker correlation with age (IC4$_{rep}$: F=37, p=1x10^{-5}; IC8$_{rep}$: F=14, p=0.012). The same pattern of age-by-group interaction observed in IC5 and IC9 of the discovery sample was also seen in IC4$_{rep}$ and IC8$_{rep}$ with the AD group showing a negative (t=-4.5, d=-0.67, p=9.8x10^{-6}, uncorrected) and positive (t=5.9, d=0.87, p=1.8x10^{-8}, uncorrected) association with age, respectively, and the other groups showing no association (except MCI within IC4$_{rep}$: t=-4.1, d=-0.41, p=5.9x10^{-5}, uncorrected).

3.3 Independent components capture structural variation due to differences in scanner hardware

IC4, IC8, IC12 and IC19 showed large effect of head coil (IC4: F=36.9, IC8: F=218.8, IC12: F= 144.9, IC19: F=60.5, p < 0.001, Fig. S5). Visual inspection of the spatial maps (Fig. S6) confirmed that IC8,12,19 appeared to be non-anatomical and were removed from further analyses.

3.4 Multivariate analyses on LICA subject loadings

3.4.1 Pairwise group classification

Table 2 shows the results from the group classifications. In the discovery sample, classification performance was high for AD *versus* MCI (AUC=0.80, sensitivity = 0.73, specificity = 74), AD *versus* SCI (AUC=0.85, sensitivity = 0.79, specificity = 76), while being not different than chance for MCI *versus* SCI (AUC=0.53, sensitivity = 0.52, specificity = 0.46, p>0.05, permutation testing). The classifier showed very good performance for classification of AD *versus* HCO (AUC=0.87, sensitivity = 0.82, specificity = 0.76), and chance level accuracy for classification of MCI or SCI *versus* HCO. Across all classification pairs, relatively balanced sensitivity and specificity were observed.

When used alone, IC9 showed higher and comparable classification performance compared to IC5 and IC0, respectively. In particular, IC0 yielded an AUC of 0.73, 0.77 and 0.84 for AD *versus* MCI, SCI, and HCO, respectively. IC5 yielded an AUC of 0.62, 0.69 and 0.57. IC9 yielded higher performance than IC5 in all cases and higher than IC0 for AD *versus* MCI, SCI (AUC=0.77, 0.79) and lower for AD *versus* HCO (AUC=0.83) (all tested via bootstrapping with stratified sampling with replacement and 10000 iterations, p<0.0001).

In the replication sample, classification results (AUC = 0.93, 0.71, 0.70 for AD *versus* HC, AD *versus* MCI, MCI *versus* HC, respectively, Table 2) showed a similar performance as obtained in the discovery sample, although slightly higher for AD *versus* HC, MCI *versus* HC and lower for AD *versus* MCI.

3.4.2 Age prediction

Fig. S8B shows the prediction performance (R^2) for the multivariate age predictor of the early, late, and full age ranges (R^2=0.42, 0.62, 0.77, respectively). The performance was lower for narrowed age ranges. Fig. 4 shows the relevant feature importance. IC9 showed a weak contribution to the age prediction across all age ranges, whereas IC0 was the most

important feature for all age classifiers. The MTL pattern (IC5) appeared to show different involvement at different age ranges. Specifically, whereas it showed strong contribution in early aging (β=0.20, rank=4) prediction, there was no contribution from this component at late aging (β ≅ 0), indicating that this component is not sensitive to age at the elderly age ranges. Similarly, IC3 and IC7 (reflecting global thickness and superior cortical thickness, respectively, Fig. S7) were sensitive to early but not late aging. Conversely, IC1 (reflecting global surface area), IC14 (ventricular enlargement) and IC27 (middle frontal surface area and GMD) (Fig. S7) were among the most important features in late but not early aging.

When used alone, whereas IC0 yielded moderate age prediction performance (R^2=0.2, 0.49, 0.65 for early, late and full ranges, respectively) compared to using all LICA feature, IC5 showed modest and no predictive power at early (R^2=0.09) and late (R^2=0.002) ranges, respectively, and IC9 showed no predictive power (R^2=0.003, 0.03).

3.4.3 Associations with cognition and MMSE prediction

IC0, 5, and 9 correlated positively with MMSE within AD (IC0: partial r = 0.33, p=9x10^{-5}; IC5: r = 0.24, p=0.0045; IC9: r = 0.39, p=1.59x10^{-6}, Fig. S9), indicating that individual differences in these components in AD patients are associated with disease severity. IC5 also correlated positively with MMSE within SCI (r=0.44, p=0.0068, Fig. S9). No significant association was found within MCI. In terms of prediction, an R^2 of 0.12 (r=0.35, p <0.05, permutation) was observed between the predicted and observed MMSE within AD, in line with the univariate results. IC5 and IC9 were the most important features (Fig. 4). A poor performance was observed within MCI and SCI (R^2=0.03 and 0.01, respectively).

3.4.4 Comparisons between clinical and age classifiers

The most important features (n=5) as indicated by the standardized regression coefficients (Fig. 4) were IC0,5,9,21,37 for AD *versus* MCI, IC0,5,9,37,46 for AD *versus* SCI, IC0,7,9,21,37 for AD *versus* HCO, IC0,3,5,7,28 for early age prediction, IC0,1,14,27,40 for

late age prediction. The classifiers AD *versus* MCI and AD *versus* HCO shared 4 out of 5 most important features (IC0,5,9,37), and all three classifiers AD *versus* MCI, SCI, HCO shared 3 out of 5 most important features (IC0,9,37), whereas there is one common feature among the set of most important features between late age prediction and the group classifiers (IC0) and two between early age prediction and the group classifiers (IC0,5).

Whereas in the age classifiers, IC0, IC1, IC3 (global thickness, Fig. S7) were the main predictors (effect of other ICs were either zero and very weak), in clinical group classifications, although IC0, IC5 and IC9 were the main predictors, there was considerable effect from other components, for instance IC37 (basal ganglia GMD, Fig. S7) in AD *versus* MCI,SCI, IC7 (superior cortical thickness, Fig. S7) in AD,MCI *versus* HCO, and IC31 (temporal pole area and GMD, Fig. S7) in AD,MCI,SCI *versus* HCO.

IC0 was the most important feature for classification of AD *versus* MCI, SCI and HCO as well as for the age and MMSE classifiers. Both IC5 and IC9 showed importance in classification between AD and MCI, SCI, HCO with IC9 showing higher contribution than IC5. IC5 was important to early but not late age prediction, while but IC9 showed almost no contribution to either early or age prediction (Fig. 4).

Fig. 5 presents performance profiles when comparing among the group (AD *versus* MCI, SCI, HCO) and age classifiers (early, Age_18_45 age prediction, and late, Age_45_90 age prediction). When excluding features based on its own feature importance, Age_18_45 dropped to a low performance (R^2<0.1) after excluding the first five most important features, while Age_45_90 after the first one. The group classifier dropped the performance to chance level (0.5) after excluding a larger number of important features (approximately 12, 7 and 8 features for AD *versus* MCI, SCI and HCO, respectively). The group classifiers, when referring to each other, showed similar profiles, whereas highly different performance profiles were observed, mostly reflecting a drop after IC0 was excluded, when referring to

the age classifiers. Similarly, the performance profile of the age classifiers showed considerable difference when referring to the group classifiers compared to when referring to their own feature importance rankings. Within the age classifiers, the performance profile of the early age classifier changed considerably when referring to late age classifier, reflecting a drop in accuracy after IC0 was dropped and then stayed rather stable. Unlike the early age classifier, the performance of the late age classifier dropped significantly after IC0 was excluded.

4 Discussion

Using a data-driven multivariate approach, we found distinct brain patterns that are sensitive to clinical status and useful for AD classification. The patterns included both age-related and disease-specific modes of gray matter variation related to AD, which were consistent across the discovery and replication samples. A global thickness and gray matter density pattern represents a common shared feature between aging and AD. Disease-specific features included an anterior-to-posterior thickness and gray matter gradient, and medial temporal lobe atrophy in late age. The reported multivariate patterns also showed moderate predictive value for cognitive status in AD patients. Using machine learning, we attempted to disentangle dementia from normal aging and report evidence of differential involvement of morphological patterns useful for age prediction during different parts of the adult lifespan and clinical status classification, supporting the hypothesis of different biological mechanisms underlying normal brain aging and AD-related neurodegeneration.

4.1 Patterns sensitive to clinical diagnosis

The global IC0 was both sensitive to age and diagnosis suggesting common mechanisms. IC0 was characterized by a well-known pattern of widespread cortical thinning (Douaud et al., 2014; Groves et al., 2011). A bidirectional pattern of GMD including anterior temporal

increases was also seen. The nature of the bidirectional GMD-pattern is unknown, but the finding suggests a regional variation in the relationship between thickness and GMD, and may be related to variable age-differences in tissue intensity and contrast (Salat et al., 2009; Westlye et al., 2010). More research is needed to understand the biological relevance of GMD.

IC5 reflected a characteristic morphological pattern primarily encompassing the MTL, implicating coordinated cortical thickness and GMD reductions in the entorhinal cortex, parahippocampus, hippocampus and the insular cortex in AD. The neuroanatomical distribution of this component is consistent with one of the most established neuroimaging biomarkers of AD, and the curvilinear age trajectories (Ostby et al., 2009; Walhovd et al., 2011) and the moderate correlation of the subject weights with hippocampus volume suggest that this component is partly reflecting the volume of the hippocampal structure along with other and coordinated MTL structural differences. Although numerous studies have implicated the MTL as a sensitive marker in differentiating AD from MCI and HC (Dickerson et al., 2009; Duara et al., 2008; Frisoni et al., 2010), the present results support the notion that MTL atrophy is not very specific for AD (Fjell et al., 2013a; Likeman et al., 2005).

While being discriminative, the MTL component (IC5, Fig. 2) was outperformed by IC0 and IC9 on its discriminative power in differentiating between AD and MCI, SCI and HCO. Singh et al. observed widespread cortical thinning with significant extension into the lateral temporal lobe associated with disease progression from MCI to AD (Singh et al., 2006). The associations of IC0 and IC9 are consistent with these results, implicating global cortical thinning and additional thinning at the lateral temporal lobe. Whereas IC0 showed strong relations with age across the entire age span included and IC5 across the early age span, IC9 was uniquely related to diagnosis suggesting that this distinct mode of gray matter

variation is specific to AD in a manner detectable by LICA. The spatial map of IC9 consisted of a symmetrical posterior greater than anterior gradient including more cortical thinning and reduced GMD in lateral temporo-parietal and precuneus bilaterally in AD. A similar AD-specific spatial distribution involving bilateral parietal lobes has been observed (Du et al., 2007), and resembles MRI-findings of autopsy-confirmed early onset AD patients suggesting that posterior greater than anterior atrophy was the most specific for ruling in a diagnosis of AD compared with both controls and other dementia etiologies (Likeman et al., 2005). The results suggest that whereas structures such as the hippocampus and frontal lobes are sensitive to a spectrum of insults including aging, parietal atrophy may have a more selective vulnerability for AD pathology. For instance, cerebral amyloid angiopathy—commonly seen in AD-patients—has a predilection for posterior vasculature (Serrano-Pozo et al., 2011). We also speculate that the IC9 pattern might probe specific sub-entities of AD pathology, for instance associated with posterior AD or posterior cortical atrophy syndromes (Crutch et al., 2012). Further research on LICA and sub-group specificity within dementia syndromes is warranted.

The age association observed only within the AD group for IC5 and IC9 indicated that the older AD patients were more different than the control group compared to the younger AD patients in the MTL pattern (IC5), and vice versa, the younger AD patients were more different than the control group compared to the older AD patients in IC9 (Fig. 3B). This result is in line with previous findings documenting that early-onset AD had largest atrophy at the occipital and parietal lobes while late-onset AD were markedly atrophic at the hippocampus (Frisoni et al., 2007). These findings support differential mechanisms of early-*versus* late-onset AD, and future research should investigate if this pattern represents a marker of early-vs-late onset AD.

4.2 Discriminative patterns in AD classification

We observed high performance for classification between AD and HCO (AUC=0.87, sensitivity = 0.82, specificity = 0.76 in the discovery sample, and AUC=0.93, sensitivity = 0.84, specificity = 0.88 in the replication sample, Table 2), comparable to previous structural MRI studies. A comprehensive evaluation of different classifiers based on whole-brain structural MRI features (Cuingnet et al., 2011) reported a sensitivity ranging between 75% and over 81% and specificity of over 89%. Although our classification result did not outperform existing findings, we have demonstrated high clinical sensitivity of the derived LICA multivariate features at an individual level, meeting the required sensitivity of ideal biomarkers (>80%) (Weiner et al., 2013). The obtained accuracy when classifying between AD and MCI and SCI, respectively, was comparable with the accuracy obtained when classifying between AD and HCO. In line with the univariate analyses (Fig. S2), IC0,5,9 were among the most important features in these classifications. Our results correspond with previous implications of MTL regions in AD classification (Cuingnet et al., 2011), but expand previous reports by demonstrating stronger contribution of IC0 and IC9, in addition to the medial temporal lobe patterns (IC5).

Despite the generally observed graded pattern of AD<MCI<SCI across IC0,5,9, classification between MCI and SCI was at chance level, which is in line with the considerable group overlap on demographic and cognitive variables (Table 1). Whereas the profile of the present SCI group was comparable to other studies (Engvig et al., 2014; Garcia-Ptacek et al., 2014), the MCI group was younger and higher performing in terms of global cognition compared with other studies (e.g., Alladi et al., 2006; Misra et al., 2009). Although the practice of clinical diagnostics is not easily confined to strict algorithmic definitions, such as employing specific cut-offs for MMSE, using biomarkers or more detailed neuropsychological profiling with stricter criteria for MCI and SCI could have improved the

classification due to less phenotypic overlap between the two groups. Additionally, the low performance of MCI *versus* HCO and SCI *versus* HCO indicates a brain phenotypic overlap among these three groups, a possible structural correlate of the relative high—or closer to normal—cognitive functioning in the two patient groups.

Lastly, despite unbalanced group sizes in both discovery and replication samples, we obtained a fairly balanced performance between sensitivity and specificity. This demonstrates the usability of a resampling approach prior to training in a *k*-fold cross validation framework to alleviate the difference in sample size (Kuhn and Johnson, 2013; Lunardon et al., 2014), a commonly faced challenge in machine learning in neuroimaging research, in particular when utilizing clinical samples and recruitment in a clinical setting.

4.3 Age prediction and relevant patterns

The prediction model showed good performance on the entire age span (explained variance R^2=0.77, [18;90] yrs), and moderate performance for narrowed spans (R^2=0.62 for [45;90] yrs, and 0.42 for [18;45] yrs, Fig. S8). This prediction accuracy corresponds to the performance reported in other studies (Cole et al., 2015; Schnack et al., 2016), although our performance is slightly lower, which may be explained by the fact that the healthy groups in our sample cover a much wider age span compared to (Schnack et al., 2016) rendering the effective sample size per year smaller. The higher performance of late age prediction than early age prediction may be attributed to the wider age range in the late range compared to early range, and that the effects of age on the brain structure is heterogeneous and varies across the adult lifespan (Fjell et al., 2013b). IC0 (global thinning and GMD alterations) and IC1 (global surface area) were among the most informative features in most of the age predictors. This finding is consistent with previous studies which reported age-related global cortical thinning and surface area reduction across the adult lifespan (Fjell et al., 2009;

Lemaitre et al., 2012). Notably, IC5 and IC9, which showed strong and significant diagnosis effects (Fig. 1), showed very weak contribution in age prediction across all age ranges (IC9) or at the late age range (IC5), indicating disease-specific focal patterns of gray matter alterations.

Except for IC0, the set of important features for age prediction was considerably different between early and late age prediction, suggesting that features involved in younger part of the adult lifespan are not informative for predicting age in the older, and vice versa. Indeed, IC14 (reflecting ventricular enlargement, which is particularly prevalent at old age (Pfefferbaum et al., 1994), Fig. S7), and IC27 (reflecting middle frontal surface area and GMD, Fig. S7) contributed to late but not early age prediction (Fig. 4). Conversely, IC3 (reflecting global cortical thickness reduction with age at the young age range, Fig. S4, Fig. S7), IC5 (MTL, Fig. 2) and IC7 (superior cortical thickness, Fig. S7) contributed to early but not late age prediction (Fig. 4). These findings are consistent with earlier work reporting differential effects of age on cortical thickness within different age ranges, specifically cortical thickness differences appeared to be widespread at [8;30 yrs] and became more regional and less prominent at older ranges (Westlye et al., 2010). Taken together, by demonstrating largely non-overlapping feature sets involved in predicting early and late aging, our results are in line with the notion of heterogeneous and nonlinear aging of the human brain (Fjell and Walhovd, 2010), and the spatial patterns of the implicated components and age-curves of the associated subject weights may offer a novel window into distinct and statistically independent mechanisms of aging and neurodegeneration across the AD continuum.

4.4 MMSE prediction and the relevant patterns

Strongly driven by IC0,5,9 together with IC1, the MMSE classifier yielded predicted values that explained about a fifth of the total variance of the observed MMSE values in AD

patients. Our prediction accuracy within AD is in line with the results reported in (Stonnington et al., 2010), which largely implicated the medial temporal lobe (IC5) among the relevant regions. The contribution of IC0 and IC9 in this prediction also corresponds with previous studies reporting a positive association between cognitive performance (quantified by MMSE) and cortical thickness involving the frontal, temporal as well as parietal lobes in AD (Du et al., 2007; Lerch et al., 2005). In order to avoid issues of circularity related to the fact that MMSE of the AD group was considerably lower than MCI and SCI, we did not predict MMSE across AD, MCI and SCI. The low prediction accuracy within SCI may be explained by the low variance of MMSE scores within this group due to ceiling effects (Velayudhan et al., 2014).

4.5 Differences and similarities between aging and AD-related patterns

The feature importance and the performance profiles depicted in Fig. 4 and Fig. 5 showed that the effects of age is captured in a small number of distinct patterns, whereas the effect related to disease is distributed across several spatial components. Our data suggest that the modes of brain structural variation overlapping between aging and AD are globally distributed, whereas differential patterns are more localized, such as IC5 and IC9 (Fig. 2), and capture much subtler variance compared to the global patterns (Fig. S3). The considerable difference in performance profiles between the group and age classifiers (Fig. 5) provides supportive evidence of differential involvement in AD compared to age-related patterns, both for early and late aging.

Our results did not support involvement of IC9 (anterior-posterior thickness and GMD gradient pattern) in the age classifiers across the lifespan, and for IC5 (MTL pattern) at the late age range. These features however were strongly implicated in AD classification, as evidenced from both multivariate classification using all LICA features and using either IC9 or IC5. Taken together, these results indicate that the localized patterns implicating medial

temporal atrophy and thickness/GMD gradient along the anterior-posterior axis are more specific to disease mechanisms than aging.

IC0 was important in both group classification and age prediction, suggesting that aging and dementia share the same global effects, i.e. global cortical thinning and alterations in GMD. In line with our findings suggesting differential involvement of brain characteristics in aging and AD, using pattern analysis techniques, atrophy patterns associated with advanced age have been shown to only partially overlap but and notably deviate from those typically found in AD (Habes et al., 2016). Also, our findings of MTL (IC5) involvement in dementia classification corroborates previous univariate studies showing that regional MTL volume and thickness, particularly the hippocampus and entorhinal cortex distinguish well between normal controls and AD (Dickerson et al., 2009; Fjell et al., 2010), with negligible contributions from MTL surface area (Dickerson et al., 2009).

4.6 Replication analyses using ADNI

Performing the same analyses independently on a replication sample, we found a set of three components ($IC3_{rep}$, $IC4_{rep}$, $IC8_{rep}$) that strongly resembled IC0, IC5 and IC9 in the discovery sample (Fig. 2). Among these components, IC5 was nearly identical to $IC4_{rep}$, suggesting that that this MTL pattern is highly robust across different cohorts. Both IC0 and its counterpart in the replication sample ($IC3_{rep}$) showed a strong age association, although IC0 was much more strongly associated with age. This is likely explained by the much larger age variance in the discovery sample, which was captured in IC0. IC9, more strongly driven by thickness than GMD, also showed a high similarity with $IC8_{rep}$. Both of these components reflected a gradient in thickness with stronger weighting posteriorly, particularly the occipital, medial parietal and lateral temporal regions, than anteriorly. This gradient however appeared to be visually more pronounced in IC9 than in $IC8_{rep}$. Despite the large difference in age range between the two samples, the fact that $IC4_{rep,}$ $IC8_{rep}$, found in a sample consisting of well age-

matched groups, resembled IC5 and IC9 support our findings on the disease-specific characteristics of these patterns.

4.7 Strengths and limitations of the study

A strength of our study includes the inclusion of a wide spectrum of participants across both age (early – late) and disease severity (SCI, MCI, and AD) which enabled us to assess common and unique features across both the healthy adult human lifespan and the AD spectrum, as well as the interplay between AD-related and age-related patterns. Further, the use of LICA allowed us to simultaneously model global (e.g. IC0, global thickness and GMD alterations) and local (e.g. IC5, MTL atrophy, and IC9, anterior-posterior thickness gradient) independent effects across different complementary morphological measures. Since a main advantage of LICA is the ability to model shared variance across different measures, the derived components may show increased sensitivity to an effect of interest, especially in the case when the effect is subtle and present across different measures (Francx et al., 2016).

Our study, however, has a number of limitations that should be taken into account when interpreting the results. Two different head coils were used during acquisition of the discovery sample, which may lead to unwanted source of MRI signal variation. However, the effects of head coil on the estimated morphometric features were accounted for by a few number of components (IC8,12,19, Fig. S6), which captured a tremendous amount of variance in head coil, leaving the rest of the components largely unaffected or affected to a modest extent. Although replication in independent samples is needed, this result demonstrates that LICA could be a promising multivariate tool for multi-site studies, where it is highly desirable to combine data from different scanners in the same unified analysis framework. Secondly, although the study was designed to assess gray matter differences only, inspection of the FSL-VBM GM-segmentation revealed some inclusion of underlying WM across the cortex likely due to age-related blurring of tissue borders and the probabilistic

threshold set for tissue separation. We have recently demonstrated the utility of LICA for diffusion MRI metrics in an overlapping dataset (Doan et al., 2017) and future studies should assess the benefits of combining features across imaging modalities (morphometry, diffusion MRI, functional MRI, etc.).

In this study, we used a model order of 50 for LICA decomposition. Our classification and prediction results indicate stable performance across a range of model orders and hierarchical clustering results favor this choice of model order. However, there is generally no optimal model order and future studies should assess the use of different model orders.

Another limitation is that the AD group in the discovery sample was not very well age-matched with the other elderly groups, making it difficult to completely rule out the effect of age in the group comparisons, although age was included as covariate in all univariate analyses. Since advanced age is the single most important risk factor for AD, the slightly higher age in the AD group compared to the MCI and SCI group is not surprising, in particular considering the clinical nature of the study. Whereas the results from the replication sample strongly support that the results from the discovery sample are not simply explained by age-differences between groups, follow-up studies are needed to assess the value of the identified brain patterns for predicting clinical conversion in the MCI and SCI patients, which has important clinical implications. Lastly, whereas the diagnostic workup was performed by two experienced physicians according to research criteria and following a comprehensive and standardized protocol (Braekhus et al., 2011), the clinical and biological phenotyping (e.g., amyloid and tau status, APOE4) in the current discovery samples was limited. Further studies are needed to test for associations with a wider range of clinical and cognitive phenotypes, including amyloid status, genetic risk, episodic memory etc.

Conclusively, by means of data-driven analysis, we have reported distinct and statistically independent multivariate MRI-based brain patterns across a naturalistic memory

clinic patient sample with increasing degree of cognitive impairment showing high sensitivity to clinical status. In addition to global reductions in estimated cortical thickness and surface area, which were isolated in independent components capturing large data variance, the anatomical distribution of the other clinically sensitive components, capturing subtler variance, are in line with known pathophysiological properties of AD, reflecting co-occurring thickness and GMD reductions encompassing MTL regions including the hippocampus, lateral temporal, precuneus and posterior cingulate cortex, as well as a pattern of anterior-posterior thickness and GMD gradient. In addition to its clinical sensitivity, a global thickness and GMD pattern also showed very high age prediction power, corroborating the extant evidence of reduced apparent cortical thickness throughout the adult lifespan. These patterns are consistent across independent samples. The current findings expand previous knowledge by suggesting that the characteristic pattern of cortical thickness and volumetric reductions in aging and across the AD spectrum are in fact reflecting the linear combination of several independent components that may represent distinct neurobiological and pathophysiological processes. Using multivariate machine learning techniques, we documented differential and specific brain characteristics involved in dementia compared to both early and late aging, and also between these different age spans. This result adds structural neuroimaging evidence to the notion that aging across the adult lifespan is heterogeneous and non-linear, and that AD reflects accelerated aging alongside disease-specific effects, supporting the utilization of advanced structural MRI for early detection in a clinical dementia setting.

5 Acknowledgements

The work was supported by the European Commission's 7th Framework Programme (#602450, IMAGEMEND), Research Council of Norway (213837, 223273, 204966/F20), the

South-Eastern Norway Regional Health Authority (2013123, 2014097, 2015073, 2016083), The Norwegian Health Association's Dementia Research Program, and KG Jebsen Foundation. We acknowledge the contribution of patient data from the Norwegian registry for persons being evaluated for cognitive symptoms in specialized care (NorCog) by the Norwegian National Advisory Unit on Ageing and Health. Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

6 References

Alladi, S., Arnold, R., Mitchell, J., Nestor, P.J., Hodges, J.R., 2006. Mild cognitive impairment: applicability of research criteria in a memory clinic and characterization of cognitive profile. Psychol Med 36, 507-515.

Andersson, J.L., Jenkinson, M., Smith, S., 2007. Non-linear registration, aka Spatial normalisation FMRIB technical report TR07JA2. FMRIB Analysis Group of the University of Oxford 2.

Braekhus, A., Ulstein, I., Wyller, T.B., Engedal, K., 2011. The Memory Clinic--outpatient assessment when dementia is suspected. Tidsskr Nor Laegeforen 131, 2254-2257.

Brodersen, K.H., Ong, C.S., Stephan, K.E., Buhmann, J.M., 2010. The balanced accuracy and its posterior distribution. Pattern recognition (ICPR), 2010 20th international conference on. IEEE, pp. 3121-3124.

Buckner, R.L., Snyder, A.Z., Shannon, B.J., LaRossa, G., Sachs, R., Fotenos, A.F., Sheline, Y.I., Klunk, W.E., Mathis, C.A., Morris, J.C., 2005. Molecular, structural, and functional characterization of Alzheimer's disease: evidence for a relationship between default activity, amyloid, and memory. The Journal of Neuroscience 25, 7709-7717.

Cai, W., Ryali, S., Chen, T., Li, C.-S.R., Menon, V., 2014. Dissociable roles of right inferior frontal cortex and anterior insula in inhibitory control: evidence from intrinsic and task-related functional parcellation, connectivity, and response profile analyses across multiple datasets. The Journal of Neuroscience 34, 14652-14667.

Chen, K., Reiman, E.M., Huan, Z., Caselli, R.J., Bandy, D., Ayutyanont, N., Alexander, G.E., 2009. Linking functional and structural brain images with multivariate network analyses: a novel application of the partial least square method. Neuroimage 47, 602-610.

Cole, J.H., Leech, R., Sharp, D.J., 2015. Prediction of brain age suggests accelerated atrophy after traumatic brain injury. Annals of neurology 77, 571-581.

Crutch, S.J., Lehmann, M., Schott, J.M., Rabinovici, G.D., Rossor, M.N., Fox, N.C., 2012. Posterior cortical atrophy. Lancet Neurol 11, 170-178.

Cuingnet, R., Gerardin, E., Tessieras, J., Auzias, G., Lehéricy, S., Habert, M.-O., Chupin, M., Benali, H., Colliot, O., Initiative, A.s.D.N., 2011. Automatic classification of patients with Alzheimer's disease from structural MRI: a comparison of ten methods using the ADNI database. NeuroImage 56, 766-781.

Dale, A.M., Fischl, B., Sereno, M.I., 1999. Cortical surface-based analysis: I. Segmentation and surface reconstruction. NeuroImage 9, 179-194.

Desikan, R.S., Cabral, H.J., Hess, C.P., Dillon, W.P., Glastonbury, C.M., Weiner, M.W., Schmansky, N.J., Greve, D.N., Salat, D.H., Buckner, R.L., 2009. Automated MRI measures identify individuals with mild cognitive impairment and Alzheimer's disease. Brain, awp123.

Dickerson, B.C., Feczko, E., Augustinack, J.C., Pacheco, J., Morris, J.C., Fischl, B., Buckner, R.L., 2009. Differential effects of aging and Alzheimer's disease on medial temporal lobe cortical thickness and surface area. Neurobiology of aging 30, 432-440.

Doan, N.T., Engvig, A., Persson, K., Alnaes, D., Kaufmann, T., Rokicki, J., Cordova-Palomera, A., Moberget, T., Braekhus, A., Barca, M.L., Engedal, K., Andreassen, O.A., Selbaek, G., Westlye, L.T., 2017. Dissociable diffusion MRI patterns of white matter microstructure and connectivity in Alzheimer's disease spectrum. Scientific Reports 7, 45131.

Doan, N.T., Kaufmann, T., Bettella, F., Jørgensen, K.N., Brandt, C.L., Moberget, T., Alnæs, D., Ueland, T., Douaud, G., Duff, E., Djurovic, S., Melle, I., Agartz, I., Andreassen, O.A., Westlye, L.T., under review. Distinct multivariate brain morphological patterns and their added predictive value with cognitive and polygenic risk scores in mental disorders. Neuroimage:Clinical.

Dørum, E., Alnæs, D., Kaufmann, T., Richard, G., Lund, M., Tønnesen, S., Sneve, M., Mathiesen, N., Rustan, Ø., Gjertsen, Ø., Vatn, S., Fure, B., Andreassen, O., Nordvik , J., Westlye, L., 2016. Age-related differences in brain network activation and co-activation during multiple object tracking. Brain and Behavior In press.

Dørum, E.S., Kaufmann, T., Alnæs, D., Andreassen, O.A., Richard, G., Kolskår, K.K., Nordvik, J.E., Westlye, L.T., 2017. Increased sensitivity to age-related differences in brain functional connectivity during continuous multiple object tracking compared to resting-state. NeuroImage 148, 364-372.

Douaud, G., Groves, A.R., Tamnes, C.K., Westlye, L.T., Duff, E.P., Engvig, A., Walhovd, K.B., James, A., Gass, A., Monsch, A.U., Matthews, P.M., Fjell, A.M., Smith, S.M., Johansen-Berg, H., 2014. A common brain network links development, aging, and vulnerability to disease. Proc Natl Acad Sci U S A 111, 17648-17653.

Douaud, G., Smith, S., Jenkinson, M., Behrens, T., Johansen-Berg, H., Vickers, J., James, S., Voets, N., Watkins, K., Matthews, P.M., 2007. Anatomically related grey and white matter abnormalities in adolescent-onset schizophrenia. Brain 130, 2375-2386.

Du, A.-T., Schuff, N., Kramer, J.H., Rosen, H.J., Gorno-Tempini, M.L., Rankin, K., Miller, B.L., Weiner, M.W., 2007. Different regional patterns of cortical thinning in Alzheimer's disease and frontotemporal dementia. Brain 130, 1159-1166.

Duara, R., Loewenstein, D.A., Potter, E., Appel, J., Greig, M.T., Urs, R., Shen, Q., Raj, A., Small, B., Barker, W., Schofield, E., Wu, Y., Potter, H., 2008. Medial temporal lobe atrophy on MRI scans and the diagnosis of Alzheimer disease. Neurology 71, 1986-1992.

Dukart, J., Schroeter, M.L., Mueller, K., Initiative, A.s.D.N., 2011. Age correction in dementia–matching to a healthy brain. PloS one 6, e22193.

Engvig, A., Fjell, A.M., Westlye, L.T., Skaane, N.V., Dale, A.M., Holland, D., Due-Tonnessen, P., Sundseth, O., Walhovd, K.B., 2014. Effects of cognitive training on gray matter volumes in memory clinic patients with subjective memory impairment. J Alzheimers Dis 41, 779-791.

Farris, J.S., 1969. On the cophenetic correlation coefficient. Systematic Biology 18, 279-285.

Fischl, B., Dale, A.M., 2000. Measuring the thickness of the human cerebral cortex from magnetic resonance images. Proceedings of the National Academy of Sciences 97, 11050-11055.

Fjell, A.M., McEvoy, L., Holland, D., Dale, A.M., Walhovd, K.B., Alzheimer's Disease Neuroimaging, I., 2013a. Brain changes in older adults at very low risk for Alzheimer's disease. J Neurosci 33, 8237-8242.

Fjell, A.M., Walhovd, K.B., 2010. Structural brain changes in aging: courses, causes and cognitive consequences. Rev Neurosci 21, 187-221.

Fjell, A.M., Walhovd, K.B., Fennema-Notestine, C., McEvoy, L.K., Hagler, D.J., Holland, D., Brewer, J.B., Dale, A.M., Alzheimer's Disease Neuroimaging, I., 2010. CSF biomarkers in prediction of cerebral and clinical change in mild cognitive impairment and Alzheimer's disease. J Neurosci 30, 2088-2101.

Fjell, A.M., Westlye, L.T., Amlien, I., Espeseth, T., Reinvang, I., Raz, N., Agartz, I., Salat, D.H., Greve, D.N., Fischl, B., Dale, A.M., Walhovd, K.B., 2009. High consistency of regional cortical thinning in aging across multiple samples. Cereb Cortex 19, 2001-2012.

Fjell, A.M., Westlye, L.T., Grydeland, H., Amlien, I., Espeseth, T., Reinvang, I., Raz, N., Dale, A.M., Walhovd, K.B., Alzheimer Disease Neuroimaging, I., 2014. Accelerating cortical thinning: unique to dementia or universal in aging? Cereb Cortex 24, 919-934.

Fjell, A.M., Westlye, L.T., Grydeland, H., Amlien, I., Espeseth, T., Reinvang, I., Raz, N., Holland, D., Dale, A.M., Walhovd, K.B., 2013b. Critical ages in the life course of the adult brain: nonlinear subcortical aging. Neurobiology of aging 34, 2239-2247.

Francx, W., Llera, A., Mennes, M., Zwiers, M.P., Faraone, S.V., Oosterlaan, J., Heslenfeld, D., Hoekstra, P.J., Hartman, C.A., Franke, B., 2016. Integrated analysis of gray and white matter alterations in attention-deficit/hyperactivity disorder. NeuroImage: Clinical 11, 357-367.

Friedman, J., Hastie, T., Tibshirani, R., 2010. Regularization paths for generalized linear models via coordinate descent. Journal of statistical software 33, 1.

Frisoni, G.B., Fox, N.C., Jack, C.R., Jr., Scheltens, P., Thompson, P.M., 2010. The clinical use of structural MRI in Alzheimer disease. Nat Rev Neurol 6, 67-77.

Frisoni, G.B., Pievani, M., Testa, C., Sabattoli, F., Bresciani, L., Bonetti, M., Beltramello, A., Hayashi, K.M., Toga, A.W., Thompson, P.M., 2007. The topography of grey matter involvement in early and late onset Alzheimer's disease. Brain 130, 720-730.

Garcia-Ptacek, S., Cavallin, L., Kareholt, I., Kramberger, M.G., Winblad, B., Jelic, V., Eriksdotter, M., 2014. Subjective cognitive impairment subjects in our clinical practice. Dement Geriatr Cogn Dis Extra 4, 419-430.

Good, C.D., Johnsrude, I.S., Ashburner, J., Henson, R.N., Fristen, K., Frackowiak, R.S., 2002. A voxel-based morphometric study of ageing in 465 normal adult human brains. Biomedical Imaging, 2002. 5th IEEE EMBS International Summer School on. IEEE, p. 16 pp.

Groves, A.R., Beckmann, C.F., Smith, S.M., Woolrich, M.W., 2011. Linked independent component analysis for multimodal data fusion. Neuroimage 54, 2198-2217.

Groves, A.R., Smith, S.M., Fjell, A.M., Tamnes, C.K., Walhovd, K.B., Douaud, G., Woolrich, M.W., Westlye, L.T., 2012. Benefits of multi-modal fusion analysis on a large-scale dataset: life-span patterns of inter-subject variability in cortical morphometry and white matter microstructure. Neuroimage 63, 365-380.

Habes, M., Janowitz, D., Erus, G., Toledo, J., Resnick, S., Doshi, J., Van der Auwera, S., Wittfeld, K., Hegenscheid, K., Hosten, N., 2016. Advanced brain aging: relationship with epidemiologic and genetic risk factors, and overlap with Alzheimer disease atrophy patterns. Translational psychiatry 6, e775.

Herrup, K., 2010. Reimagining Alzheimer's disease--an age-based hypothesis. J Neurosci 30, 16755-16762.

Herrup, K., 2015. The case for rejecting the amyloid cascade hypothesis. Nature neuroscience, 794-799.

Hogstrom, L.J., Westlye, L.T., Walhovd, K.B., Fjell, A.M., 2013. The structure of the cerebral cortex across adult life: age-related patterns of surface area, thickness, and gyrification. Cereb Cortex 23, 2521-2530.

Hutton, C., Draganski, B., Ashburner, J., Weiskopf, N., 2009. A comparison between voxel-based cortical thickness and voxel-based morphometry in normal aging. NeuroImage 48, 371-380.

Jack, C.R., Bernstein, M.A., Fox, N.C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P.J., L Whitwell, J., Ward, C., 2008. The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. Journal of magnetic resonance imaging 27, 685-691.

James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. An introduction to statistical learning: with Applications in R. New York: springer.

Klöppel, S., Abdulkadir, A., Jack, C.R., Koutsouleris, N., Mourão-Miranda, J., Vemuri, P., 2012. Diagnostic neuroimaging across diseases. NeuroImage 61, 457-463.

Koutsouleris, N., Meisenzahl, E.M., Borgwardt, S., Riecher-Rössler, A., Frodl, T., Kambeitz, J., Köhler, Y., Falkai, P., Möller, H.-J., Reiser, M., 2015. Individualized differential diagnosis of schizophrenia and mood disorders using neuroanatomical biomarkers. Brain 138, 2059-2073.

Kuhn, M., 2008. Caret package. Journal of Statistical Software 28.

Kuhn, M., Johnson, K., 2013. Applied predictive modeling. Springer.

Lemaitre, H., Goldman, A.L., Sambataro, F., Verchinski, B.A., Meyer-Lindenberg, A., Weinberger, D.R., Mattay, V.S., 2012. Normal age-related brain morphometric changes: nonuniformity across cortical thickness, surface area and gray matter volume? Neurobiology of aging 33, 617. e611-617. e619.

Lerch, J.P., Pruessner, J.C., Zijdenbos, A., Hampel, H., Teipel, S.J., Evans, A.C., 2005. Focal decline of cortical thickness in Alzheimer's disease identified by computational neuroanatomy. Cerebral cortex 15, 995-1001.

Likeman, M., Anderson, V.M., Stevens, J.M., Waldman, A.D., Godbolt, A.K., Frost, C., Rossor, M.N., Fox, N.C., 2005. Visual assessment of atrophy on magnetic resonance imaging in the diagnosis of pathologically confirmed young-onset dementias. Arch Neurol 62, 1410-1415.

Lowe, V.J., Peller, P.J., Weigand, S.D., Montoya Quintero, C., Tosakulwong, N., Vemuri, P., Senjem, M.L., Jordan, L., Jack, C.R., Jr., Knopman, D., Petersen, R.C., 2013. Application of the National Institute on Aging-Alzheimer's Association AD criteria to ADNI. Neurology 80, 2130-2137.

Lunardon, N., Menardi, G., Torelli, N., 2014. ROSE: A Package for Binary Imbalanced Learning. A peer-reviewed, open-access publication of the R Foundation for Statistical Computing, 79.

Misra, C., Fan, Y., Davatzikos, C., 2009. Baseline and longitudinal patterns of brain atrophy in MCI patients, and their use in prediction of short-term conversion to AD: results from ADNI. NeuroImage 44, 1415-1422.

Ostby, Y., Tamnes, C.K., Fjell, A.M., Westlye, L.T., Due-Tonnessen, P., Walhovd, K.B., 2009. Heterogeneity in subcortical brain development: A structural magnetic resonance imaging study of brain maturation from 8 to 30 years. J Neurosci 29, 11772-11782.

Palaniyappan, L., Liddle, P.F., 2012. Differential effects of surface area, gyrification and cortical thickness on voxel based morphometric deficits in schizophrenia. Neuroimage 60, 693-699.

Panizzon, M.S., Fennema-Notestine, C., Eyler, L.T., Jernigan, T.L., Prom-Wormley, E., Neale, M., Jacobson, K., Lyons, M.J., Grant, M.D., Franz, C.E., Xian, H., Tsuang, M., Fischl, B., Seidman, L., Dale, A., Kremen, W.S., 2009. Distinct genetic influences on cortical surface area and cortical thickness. Cerebral Cortex 19, 2728-2735.

Petersen, R.C., Aisen, P., Beckett, L.A., Donohue, M., Gamst, A., Harvey, D.J., Jack, C., Jagust, W., Shaw, L., Toga, A., 2010. Alzheimer's disease Neuroimaging Initiative (ADNI) clinical characterization. Neurology 74, 201-209.

Pfefferbaum, A., Mathalon, D.H., Sullivan, E.V., Rawles, J.M., Zipursky, R.B., Lim, K.O., 1994. A quantitative magnetic resonance imaging study of changes in brain morphology from infancy to late adulthood. Archives of neurology 51, 874-887.

Sabuncu, M.R., Konukoglu, E., Initiative, A.s.D.N., 2015. Clinical prediction from structural brain MRI scans: a large-scale empirical study. Neuroinformatics 13, 31-46.

Salat, D.H., Lee, S.Y., van der Kouwe, A.J., Greve, D.N., Fischl, B., Rosas, H.D., 2009. Age-associated alterations in cortical gray and white matter signal intensity and gray to white matter contrast. NeuroImage 48, 21-28.

Schnack, H.G., van Haren, N.E., Nieuwenhuis, M., Pol, H.E.H., Cahn, W., Kahn, R.S., 2016. Accelerated Brain Aging in Schizophrenia: A Longitudinal Pattern Recognition Study. American Journal of Psychiatry.

Serrano-Pozo, A., Frosch, M.P., Masliah, E., Hyman, B.T., 2011. Neuropathological alterations in Alzheimer disease. Cold Spring Harb Perspect Med 1, a006189.

Singh, V., Chertkow, H., Lerch, J.P., Evans, A.C., Dorr, A.E., Kabani, N.J., 2006. Spatial patterns of cortical thinning in mild cognitive impairment and Alzheimer's disease. Brain 129, 2885-2893.

Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E., Johansen-Berg, H., Bannister, P.R., De Luca, M., Drobnjak, I., Flitney, D.E., 2004. Advances in functional and structural MR image analysis and implementation as FSL. NeuroImage 23, S208-S219.

Stonnington, C.M., Chu, C., Klöppel, S., Jack, C.R., Ashburner, J., Frackowiak, R.S., Initiative, A.D.N., 2010. Predicting clinical scores from magnetic resonance scans in Alzheimer's disease. NeuroImage 51, 1405-1413.

Sui, J., Adali, T., Yu, Q., Chen, J., Calhoun, V.D., 2012. A review of multivariate methods for multimodal fusion of brain imaging data. Journal of Neuroscience Methods 204, 68-81.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M., 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. NeuroImage 15, 273-289.

Uddin, L.Q., Supekar, K., Lynch, C.J., Khouzam, A., Phillips, J., Feinstein, C., Ryali, S., Menon, V., 2013. Salience network–based classification and prediction of symptom severity in children with autism. JAMA psychiatry 70, 869-879.

Velayudhan, L., Ryu, S.-H., Raczek, M., Philpot, M., Lindesay, J., Critchfield, M., Livingston, G., 2014. Review of brief cognitive tests for patients with suspected dementia. International psychogeriatrics 26, 1247-1262.

Wager, T.D., Atlas, L.Y., Leotti, L.A., Rilling, J.K., 2011. Predicting individual differences in placebo analgesia: contributions of brain activity during anticipation and pain experience. J Neurosci 31, 439-452.

Walhovd, K.B., Westlye, L.T., Amlien, I., Espeseth, T., Reinvang, I., Raz, N., Agartz, I., Salat, D.H., Greve, D.N., Fischl, B., Dale, A.M., Fjell, A.M., 2011. Consistent neuroanatomical age-related volume differences across multiple samples. Neurobiol Aging 32, 916-932.

Wang, W.Y., Yu, J.T., Liu, Y., Yin, R.H., Wang, H.F., Wang, J., Tan, L., Radua, J., Tan, L., 2015. Voxel-based meta-analysis of grey matter changes in Alzheimer's disease. Transl Neurodegener 4, 6.

Weiner, M.W., Aisen, P.S., Jack, C.R., Jagust, W.J., Trojanowski, J.Q., Shaw, L., Saykin, A.J., Morris, J.C., Cairns, N., Beckett, L.A., 2010. The Alzheimer's disease neuroimaging initiative: progress report and future plans. Alzheimer's & Dementia 6, 202-211. e207.

Weiner, M.W., Veitch, D.P., Aisen, P.S., Beckett, L.A., Cairns, N.J., Green, R.C., Harvey, D., Jack, C.R., Jagust, W., Liu, E., 2013. The Alzheimer's Disease Neuroimaging Initiative: a review of papers published since its inception. Alzheimer's & Dementia 9, e111-e194.

Weisberg, S., 2005. Applied linear regression. John Wiley & Sons.

Westlye, L.T., Walhovd, K.B., Dale, A.M., Bjørnerud, A., Due-Tønnessen, P., Engvig, A., Grydeland, H., Tamnes, C.K., Østby, Y., Fjell, A.M., 2010. Differentiating maturational and aging-related changes of the cerebral cortex by use of thickness and signal intensity. Neuroimage 52, 172-185.

Westman, E., Aguilar, C., Muehlboeck, J.S., Simmons, A., 2013. Regional magnetic resonance imaging measures for multivariate analysis in Alzheimer's disease and mild cognitive impairment. Brain Topogr 26, 9-23.

Westman, E., Simmons, A., Muehlboeck, J.S., Mecocci, P., Vellas, B., Tsolaki, M., Kloszewska, I., Soininen, H., Weiner, M.W., Lovestone, S., Spenger, C., Wahlund, L.O., AddNeuroMed, c., Alzheimer's Disease Neuroimaging, I., 2011. AddNeuroMed and ADNI: similar patterns of Alzheimer's atrophy and automated MRI classification accuracy in Europe and North America. NeuroImage 58, 818-828.

Wickham, H., 2009. ggplot2: elegant graphics for data analysis. Springer.

Winblad, B., Palmer, K., Kivipelto, M., Jelic, V., Fratiglioni, L., Wahlund, L.O., Nordberg, A., Backman, L., Albert, M., Almkvist, O., Arai, H., Basun, H., Blennow, K., de Leon, M., DeCarli, C., Erkinjuntti, T., Giacobini, E., Graff, C., Hardy, J., Jack, C., Jorm, A., Ritchie, K., van Duijn, C., Visser, P., Petersen, R.C., 2004. Mild cognitive impairment--beyond controversies, towards a consensus: report of the International Working Group on Mild Cognitive Impairment. J Intern Med 256, 240-246.

Winkler, A.M., Kochunov, P., Blangero, J., Almasy, L., Zilles, K., Fox, P.T., Duggirala, R., Glahn, D.C., 2010. Cortical thickness or grey matter volume? The importance of selecting the phenotype for imaging genetics studies. NeuroImage 53, 1135-1146.

World Health Organization., 1993. The ICD-10 classification of mental and behavioural disorders : diagnostic criteria for research. World Health Organization, Geneva.

	HCY (n=324)	HCO (n=31)	SCI (n=38)	MCI (n=78)	AD (n=137)	Group comparisons
Age	31.1 ± 7.5	62.8 ± 10.3	63.9 ± 9	64.2 ± 11	71.4 ± 8.3	(HCY<HCO, MCI, SCI<AD)*
Sex *N* (%) female	142 (44)	17 (55)	17 (45)	27 (35)	81 (59)	(MCI < HCO, AD; HCY < AD)*
						No difference among AD, SCI and HCO.
MMSE**	n/a	n/a	29.3 ± 0.9	28.2 ± 1.8	23 ± 4.8	(AD < MCI, SCI)*
Head coil *N* (%) 8HRBRAIN	324 (100)	31 (100)	20 (53)	44 (56)	62 (45)	(AD, MCI, SCI < HCO, HCY)*
						No difference among AD, MCI, SCI.

** missing for 1 AD, 6 MCI and 2 SCI.

Table 1. Sample characteristics. AD = Alzheimer's disease. MCI: Mild Cognitive Impairment, SCI: Subjective Cognitive Impairment, HCY: Young Controls, HCO: Elderly Controls, MMSE: Mini Mental State Examination. *Linear regression was used to compare groups for continuous variables and Chi-square test was used for categorical variables, ">" and "<" denote significant group differences ($P < .05$, Bonferroni correction) in the indicated direction.

Sample	Classification	AUC (mean ± sd)	Balanced Accuracy (mean ± sd)	Accuracy (mean ± sd)	Sensitivity (mean ± sd)	Specificity (mean ± sd)
	AD *versus* MCI	0.80 ± 0.02	0.73 ± 0.03	0.73 ± 0.03	0.73 ± 0.04	0.74 ± 0.05
	AD *versus* SCI	0.85 ± 0.02	0.77 ± 0.03	0.78 ± 0.02	0.79 ± 0.03	0.76 ± 0.05
	AD *versus* HCO	0.87 ± 0.02	0.79 ± 0.03	0.81 ± 0.02	0.82 ± 0.03	0.76 ± 0.06
	MCI *versus* SCI	0.53 ± 0.04	0.49 ± 0.04	0.50 ± 0.04	0.52 ± 0.09	0.46 ± 0.09
	MCI *versus* HCO	0.56 ± 0.06	0.54 ± 0.05	0.56 ± 0.05	0.59 ± 0.09	0.50 ± 0.10
Discovery	SCI *versus* HCO	0.53 ± 0.04	0.50 ± 0.04	0.50 ± 0.04	0.52 ± 0.10	0.47 ± 0.12
	AD *versus* MCI	0.71 ± 0.01	0.65 ± 0.02	0.66 ± 0.02	0.63 ± 0.04	0.68 ± 0.03
	MCI *versus* HC	0.70 ± 0.02	0.65 ± 0.02	0.66 ± 0.02	0.67 ± 0.03	0.63 ± 0.03
Replication	AD *versus* HC	0.93 ± 0.01	0.86 ± 0.01	0.86 ± 0.01	0.84 ± 0.03	0.88 ± 0.02

Table 2: Classification performance. sd=standard deviation.

Figure captions

Figure 1: Plot of *f*-statistics showing main effect of diagnosis and age. The dashed line represents significance threshold corrected for multiple comparisons across all components based on permutation testing.

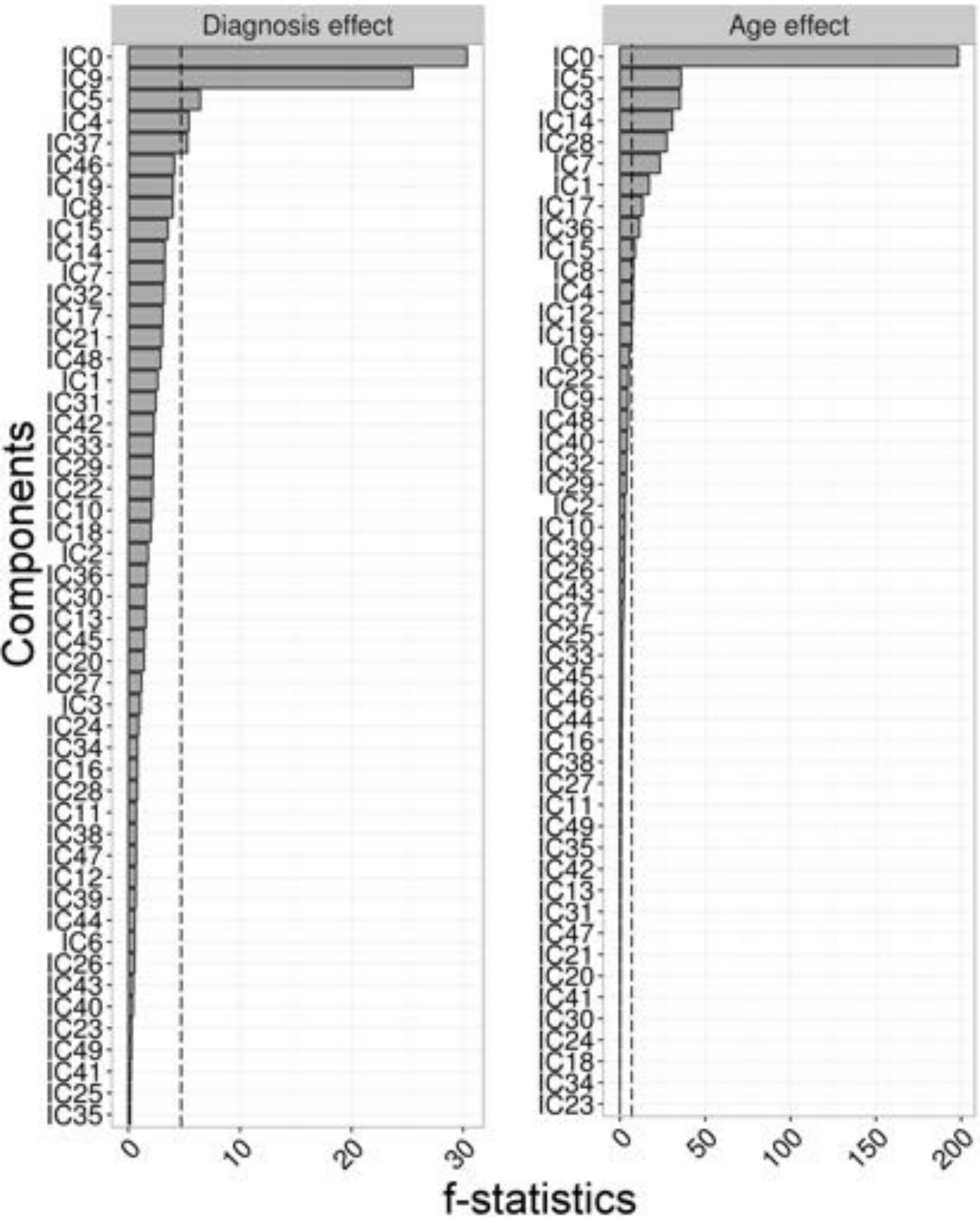
Figure 2: Spatial maps and subject loading distribution of the components showing strong diagnosis effect in the discovery (A, B) and replication samples (C, D). The spatial maps represent the thresholded z-scores ($3 < |z| < 10$). In the spatial map, the weights (in percentage) indicate the relative contribution of each measure to the component at the group level. Spatial maps of measures showing negligible contribution are not presented. All presented components are multivariate, involving both thickness and GMD. In the subject loading box plots, the box represents the 25% and 75% quantiles, the horizontal bar in the box representing the median, the diamond the mean, and the dots the outliers.

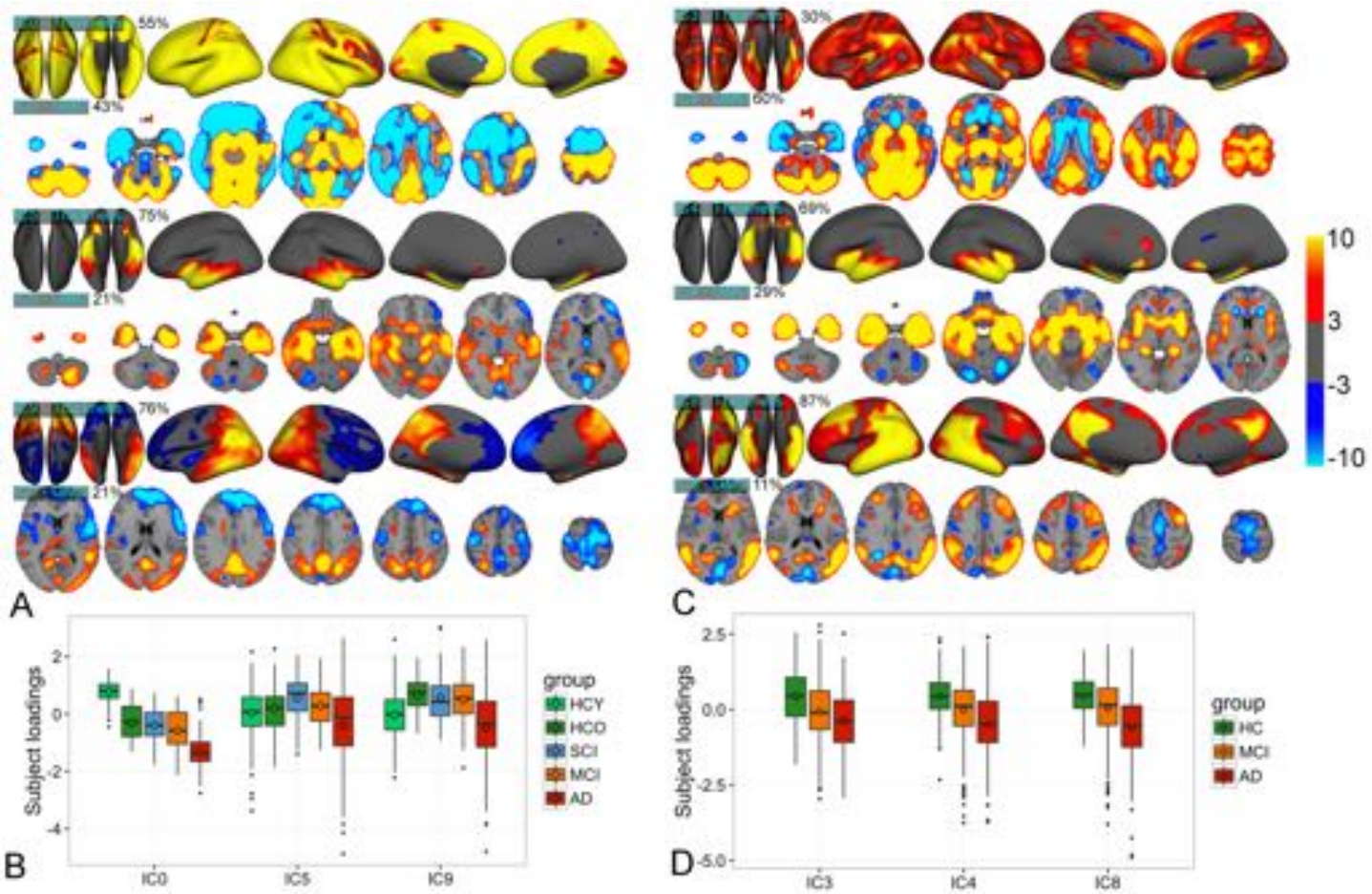
Figure 3: (A) Scatter plot of subject loadings as a function of age. Blue curves represent the LOESS fit on the combined set of HCY, HCO and SCI subjects. Red curves represent the LOESS fit using all datasets (HCY, HCO, SCI, MCI and AD). R^2=explained variance. (B) and (C) Age scatter plot within elderly healthy controls, MCI and AD of IC0,5,9 in the discovery sample (B) and IC3,4,8 in the replication sample (C). The lines represent the linear fits with groups. *rep*=replication.

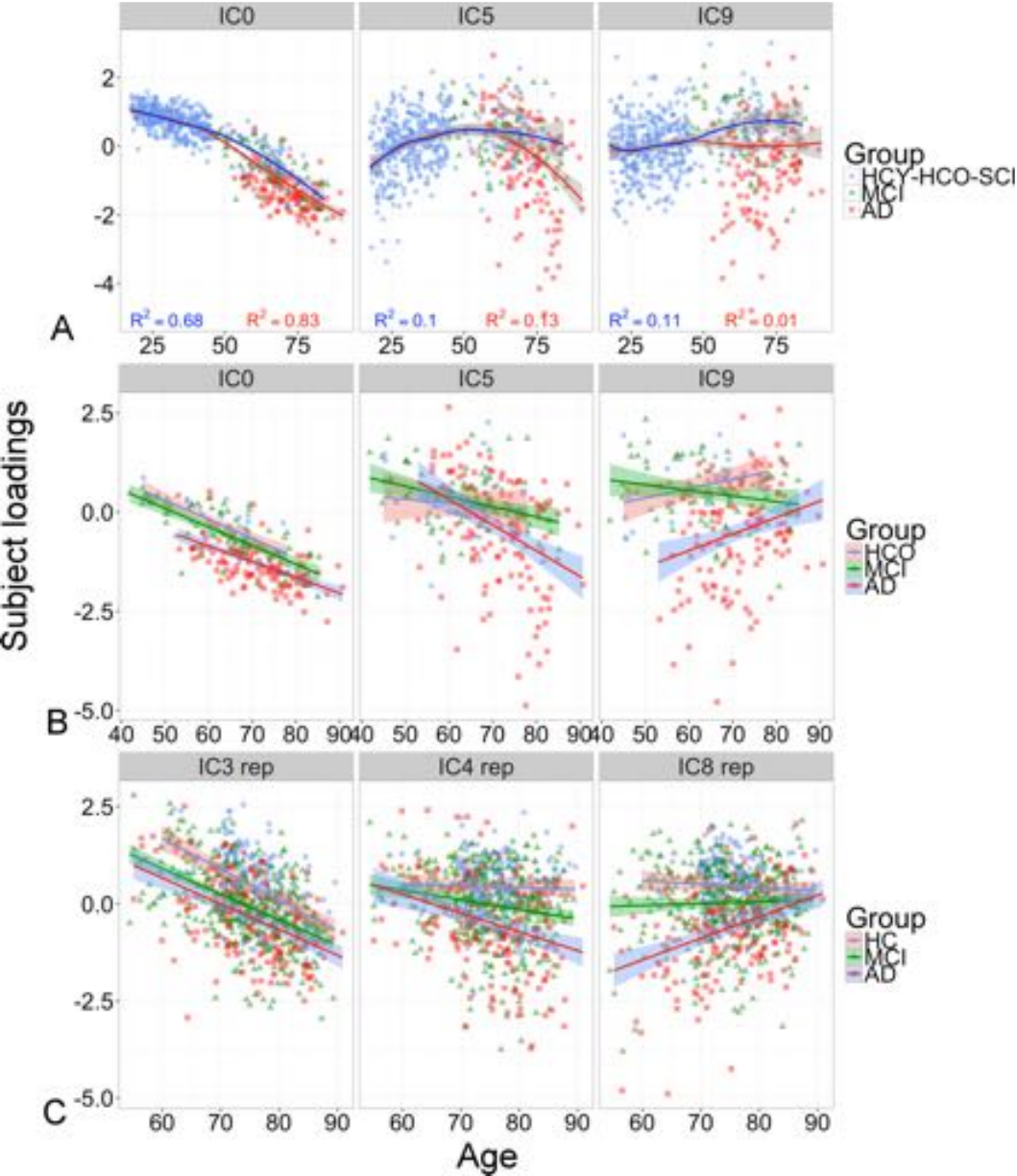
Figure 4: Feature importance quantified using the *lasso's* standardized regression coefficients. The ranking was inferred based on the magnitude of the regression coefficients. Given that most of the features showed near zero coefficients, for visualisation purpose, we focused on the first five most important features. Feature ranking of classification (MCI

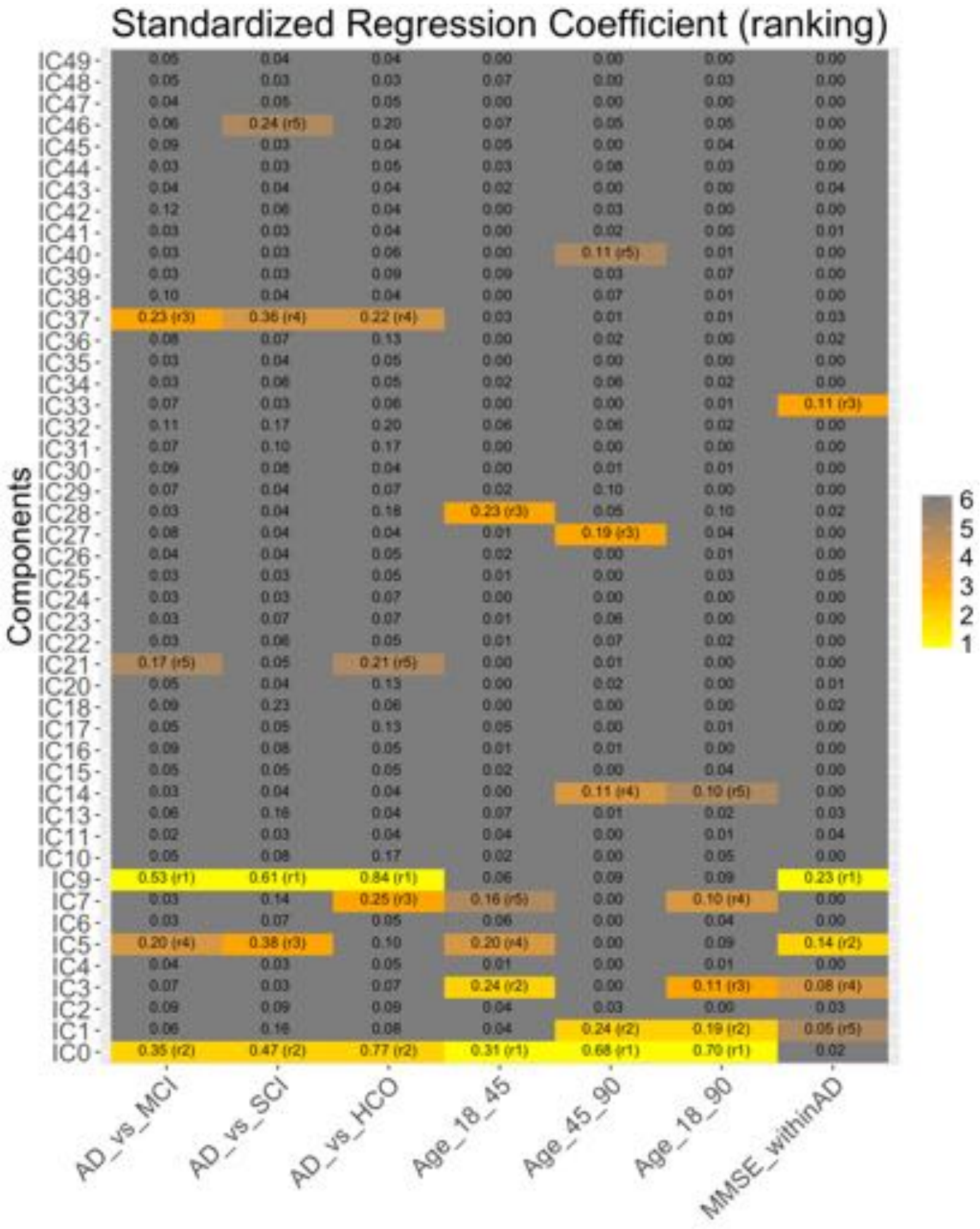
versus SCI, MCI *versus* HCO, SCI *versus* HCO) and prediction (MMSE within MCI, SCI) with low performance is not shown. The color codes for the actual feature importance rankings. r =importance ranking.

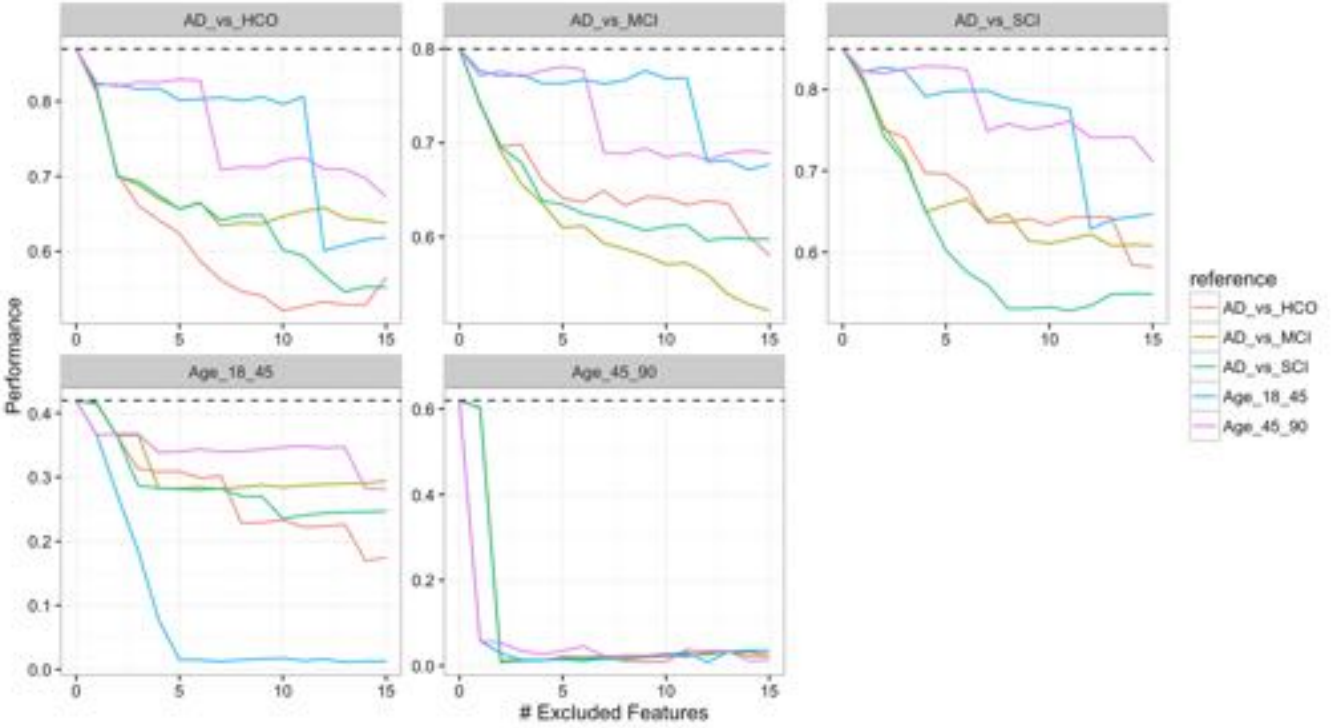
Figure 5: Performance profile of each classifier when incrementally excluding the first n^{th} most important features (n=1,2,3,…,15), referring to feature importance ranking obtained from a reference classifier to decide on the set of excluded features. Area Under the ROC Curve (AUC) and R^2 were computed for group classification and age prediction, respectively. The black dashed horizontal lines represent the performance obtained when using all LICA features.


Diagnosis effect
Age effect
Components
f-statistics
0
10
20
30
0
50
100
150
200


55%
43%
75%
21%
76%
21%
30%
60%
69%
29%
67%
11%
10
3
-3
-10
A
C
Subject loadings
group
HCY
HCO
SCI
MCI
AD
IC0
IC5
IC9
B
Subject loadings
group
HC
MCI
AD
IC3
IC4
IC8
D


IC0
IC5
IC9
A
R^2 = 0.68
R^2 = 0.83
R^2 = 0.1
R^2 = 0.13
R^2 = 0.11
R^2 = 0.01
Group
HCY-HCO-SCI
MCI
AD
Subject loadings
B
Group
HCO
MCI
AD
IC3 rep
IC4 rep
IC8 rep
C
Group
HC
MCI
AD
Age


Standardized Regression Coefficient (ranking)
Components
IC49 0.05 0.04 0.04 0.00 0.00 0.00 0.00
IC48 0.05 0.03 0.03 0.07 0.00 0.03 0.00
IC47 0.04 0.05 0.05 0.00 0.00 0.00 0.00
IC46 0.06 0.24 (r5) 0.20 0.07 0.05 0.05 0.00
IC45 0.09 0.03 0.04 0.05 0.00 0.04 0.00
IC44 0.03 0.03 0.05 0.03 0.08 0.03 0.00
IC43 0.04 0.04 0.04 0.02 0.00 0.00 0.04
IC42 0.12 0.06 0.04 0.00 0.03 0.00 0.00
IC41 0.03 0.03 0.04 0.00 0.02 0.00 0.01
IC40 0.03 0.03 0.06 0.00 0.11 (r5) 0.01 0.00
IC39 0.03 0.03 0.09 0.09 0.03 0.07 0.00
IC38 0.10 0.04 0.04 0.00 0.07 0.01 0.00
IC37 0.23 (r3) 0.36 (r4) 0.22 (r4) 0.03 0.01 0.01 0.03
IC36 0.08 0.07 0.13 0.00 0.02 0.00 0.02
IC35 0.03 0.04 0.05 0.00 0.00 0.00 0.00
IC34 0.03 0.06 0.05 0.02 0.06 0.02 0.00
IC33 0.07 0.03 0.06 0.00 0.00 0.01 0.11 (r3)
IC32 0.11 0.17 0.20 0.06 0.06 0.02 0.00
IC31 0.07 0.10 0.17 0.00 0.00 0.00 0.00
IC30 0.09 0.08 0.04 0.00 0.01 0.01 0.00
IC29 0.07 0.04 0.07 0.02 0.10 0.00 0.00
IC28 0.03 0.04 0.18 0.23 (r3) 0.05 0.10 0.02
IC27 0.08 0.04 0.04 0.01 0.19 (r3) 0.04 0.00
IC26 0.04 0.04 0.05 0.02 0.00 0.01 0.00
IC25 0.03 0.03 0.05 0.01 0.00 0.03 0.05
IC24 0.03 0.03 0.07 0.00 0.00 0.00 0.00
IC23 0.03 0.07 0.07 0.01 0.06 0.00 0.00
IC22 0.03 0.06 0.05 0.01 0.07 0.02 0.00
IC21 0.17 (r5) 0.05 0.21 (r5) 0.00 0.01 0.00 0.00
IC20 0.05 0.04 0.13 0.00 0.02 0.00 0.01
IC18 0.09 0.23 0.06 0.00 0.00 0.00 0.02
IC17 0.05 0.05 0.13 0.05 0.00 0.01 0.00
IC16 0.09 0.08 0.05 0.01 0.01 0.00 0.00
IC15 0.05 0.05 0.05 0.02 0.00 0.04 0.00
IC14 0.03 0.04 0.04 0.00 0.11 (r4) 0.10 (r5) 0.00
IC13 0.06 0.16 0.04 0.07 0.01 0.02 0.03
IC11 0.02 0.03 0.04 0.04 0.00 0.01 0.04
IC10 0.05 0.08 0.17 0.02 0.00 0.05 0.00
IC9 0.53 (r1) 0.61 (r1) 0.84 (r1) 0.06 0.09 0.09 0.23 (r1)
IC7 0.03 0.14 0.25 (r3) 0.16 (r5) 0.00 0.10 (r4) 0.00
IC6 0.03 0.07 0.05 0.06 0.00 0.04 0.00
IC5 0.20 (r4) 0.38 (r3) 0.10 0.20 (r4) 0.00 0.09 0.14 (r2)
IC4 0.04 0.03 0.05 0.01 0.00 0.01 0.00
IC3 0.07 0.03 0.07 0.24 (r2) 0.00 0.11 (r3) 0.08 (r4)
IC2 0.09 0.09 0.09 0.04 0.03 0.00 0.03
IC1 0.06 0.16 0.08 0.04 0.24 (r2) 0.19 (r2) 0.05 (r5)
IC0 0.35 (r2) 0.47 (r2) 0.77 (r2) 0.31 (r1) 0.68 (r1) 0.70 (r1) 0.02
AD_vs_MCI
AD_vs_SCI
AD_vs_HCO
Age_18_45
Age_45_90
Age_18_90
MMSE_withinAD
6
5
4
3
2
1


AD_vs_HCO
AD_vs_MCI
AD_vs_SCI
Age_18_45
Age_45_90
Performance
Excluded Features
reference
AD_vs_HCO
AD_vs_MCI
AD_vs_SCI
Age_18_45
Age_45_90

Supplemental Material

Statistical analysis

Correction for multiple testing of all GLM analyses across all LICA components was performed using permutation testing. In each iteration, we permuted each of the following variables with respect to all the others: LICA subject loadings, age, sex, group and head coil prior to performing GLM analyses. This process was repeated 10,000 times, resulting in the null distribution of the relevant *s*tatistics (Cohen's *d* or *f*) based on which the corrected *p*-value was computed.

We also assessed the significance of group classification and age, MMSE prediction based on permutation testing using the same procedure as described above. In each of the 10,000 permutation iteration, we permuted the LICA components, group label (for classification), and age, MMSE (for prediction), one with respect to all the others prior to performing classification or prediction to obtain the null distribution of the performance.

Multivariate analyses

We run the same group classification and age prediction pipeline as run on the LICA features using either the FS-VBM or PCA feature sets as input features. LICA (AUC=0.87, 0.85, 0.8 for AD vs. HCO, SCI and MCI) showed slightly higher and lower group classification performance compared to the PCA (AUC=0.83, 0.83, 0.73 for AD vs. HCO, SCI and MCI) and FS-VBM feature set (AUC=0.89, 0.90, 0.82 for AD vs. HCO, SCI and MCI), respectively (Fig. S8A). In terms of age prediction, LICA (R^2=0.42, 0.62, 0.77 for early, late and full age ranges) showed higher performance than both feature sets (PCA: R^2=0.38, 0.37, 0.71; FS-VBM: R^2=0.42, 0.44, 0.71), especially at the late age range (Fig. S8B). Fig. S8C presents the *t*-statistics of the age effect obtained by running a linear regression model on each feature in each of the feature sets using all subjects in the HCY, HCO and SCI

groups, accounting for sex. The LICA feature set showed increased sensitivity to age than the FS-VBM and PCA sets at the late or full age ranges, as suggested by the higher magnitude of the *t*-statistics. Taken together, these results suggest that LICA feature set, while being comparable to FS-VBM and PCA sets regarding group classification, it appeared to be more efficient in capturing age-related patterns, and thus better suited for the purpose of disentangling age and disease-related patterns.

LICA on thickness, area and GMD compared to ICA decompositions on each of these measures

We performed ICA on each data type (thickness, surface area or GMD maps) using the same model order (50) and compared the results with the original LICA run on all measures. Whereas the spatial maps of IC0 was seen in the corresponding IC (IC0) of single measure (thickness, GMD) ICA decompositions (subject loadings' correlation r=0.94, 0.92, respectively), IC5 LICA was presented in multiple components (maximum subject loadings' correlation with single measure components r=0.72). IC9 LICA was also only partly visible in the single ICA components (maximum correlation with single measure components r=0.58). Importantly, when comparing the effect sizes obtained using each single measure ICA and the LICA components, the LICA components showed increased sensitivity to AD (Fig. S10A). In line with the univariate results, multivariate classification results showed that the LICA components yielded superior performance compared to the single measure ICA feature sets in AD vs. MCI, SCI and HCO (Fig. S10B), with the exception that the GMD ICA features showed slightly higher performance than LICA in AD vs. HCO (AUC = 0.89 vs. 0.87). These results provide empirical evidence supporting the benefits of combining these complementary measures using LICA.

The effect of LICA model order on classification, prediction performance and model order selection

We ran LICA using a range of different model orders (d=10, 20, …, 100) and performed group classification and age prediction to assess how the performance varies with respect to *d.* The results showed that the performance is fairly stable across model orders, with the exception of the [45;90 yrs] age classifier (Fig. S11A). For this age classifier, the performance steadily increased from d=10 and became relatively stable from d=50. This result indicates that at a model order of 50 or more, the LICA patterns showed in increased sensitivity to old age compared to smaller orders.

Additionally, in an unsupervised manner, for each model order d, we performed hierrachical clustering of the subjects using *average* linkage and *Euclidean* distance. Next we used the cophenetic correlation coefficient to evaluate the resulting dendrogram. A higher value of the cophenetic coefficient represents a clustering dendrogram that better fits the data. As seen in Fig. S11B, the cophenetic coefficient significantly increased from d=10 until d=50, then started to become relatively stable (increasing with a reduced rate). This indicates that d=50 is an appropriate choice for LICA decomposition of these data.

Supplementary Figure Legends

Figure S1: [Replication sample] Main effect of diagnosis, age and sex on the LICA subject loadings. The dashed line in each subplot represents the significance threshold, corrected for the number of LICA components, obtained using permutation testing.

Figure S2: Effect size (Cohen's d) of group pairwise comparisons in (A) discovery and (B) replication samples. *, ** indicate significance levels (* $0.001<p<0.05$, ** $p<0.001$) after correcting for multiple comparison using permutation testing.

Figure S3: [Discovery sample] Percentage explained variance of all LICA components.

Figure S4: [Discovery sample] Scatter plot of subject loadings *versus* age for IC1,3,7,14. Only components where the LOESS fits across all datasets showed $R^2 \geq 0.1$ are shown. Blue curves represent the LOESS fit on the combined set of HCY, HCO and SCI subjects. Red curves represent the LOESS fit using all datasets.

Figure S5: [Discovery sample] Main effect of head coil on the components' subject loadings. The dashed line represent the permutation-based significance threshold corrected across all components.

Figure S6: [Discovery sample] Spatial maps of the components capturing strong head coil effects (IC8,12,19). The spatial maps represent the thresholded z-scores ($3 < |z| < 10$). The weights indicate the relative contribution of each measure to the component at the group level. Spatial maps of measures showing negligible contribution (weight <5%) are not presented.

Figure S7: [Discovery sample] Spatial maps of the components rather than IC0, IC5 and IC9 sensitive to AD and age. The spatial maps represent the thresholded z-scores ($3 < |z| < 10$). The weights indicate the relative contribution of each measure to the component at the group level. Spatial maps of measures showing negligible contribution (weight <5%) are not presented.

Figure S8: [Discovery sample] (A) Classification performance, (B) Age prediction performance, (C) Histogram of the *t*-statistics of the age effect on either the LICA feature (n=47), FS-VBM feature (n=262) or PCA feature (n=136) sets.

Figure S9: [Discovery sample] GLM fit within each group of subject loadings against Mini-Mental State Examination score (MMSE).

Figure S10: [Discovery sample] (A) Histogram of the Cohen's *d* and (B) Classification performance using either LICA or different ICA feature sets.

Figure S11: [Discovery sample] (A) Performance (area under ROC curve for classification and R^2 between predicted and true value for age prediction) plotted as a function of LICA model order, (B) Cophenetic correlation coefficient as a function of model order.

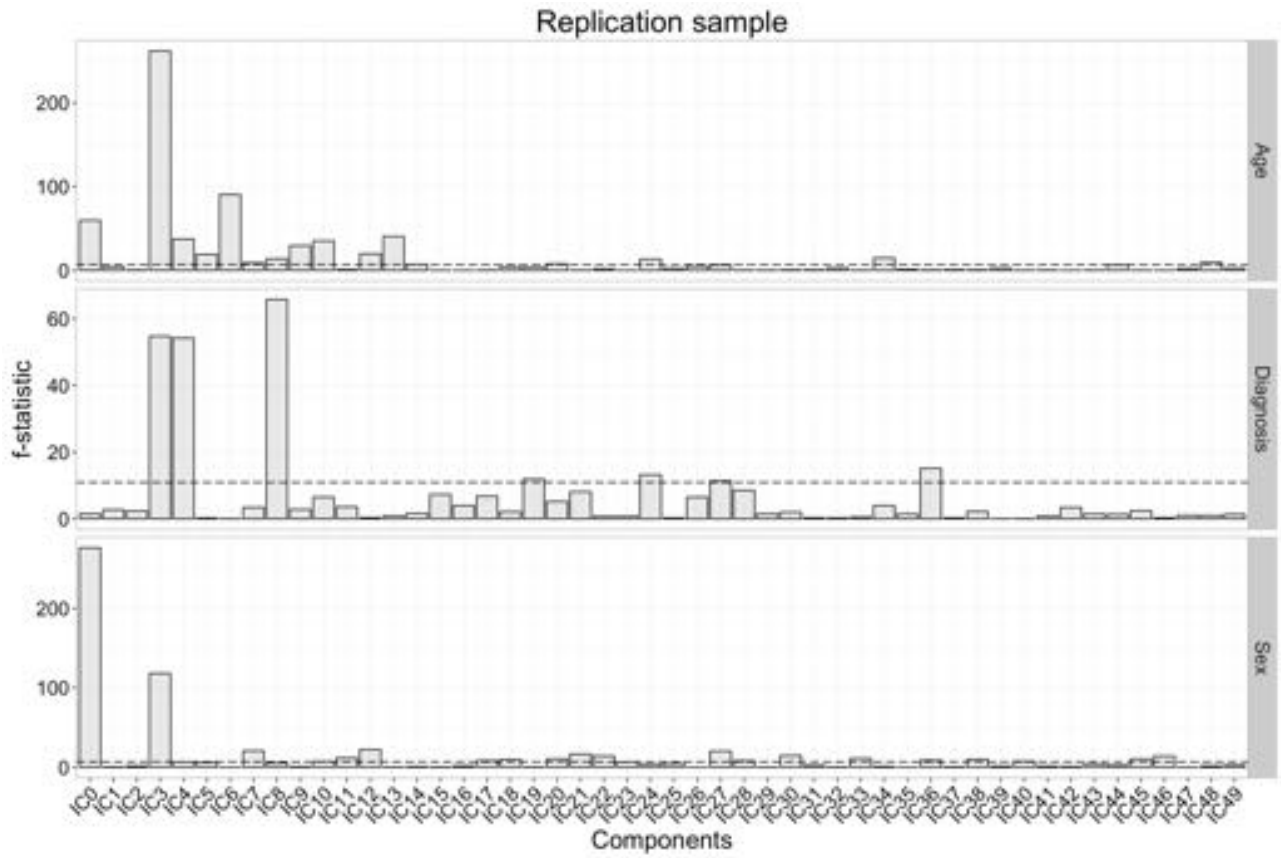


Figure S1: [Replication sample] Main effect of diagnosis, age and sex on the LICA subject loadings. The dashed line in each subplot represents the significance threshold, corrected for the number of LICA components, obtained using permutation testing.

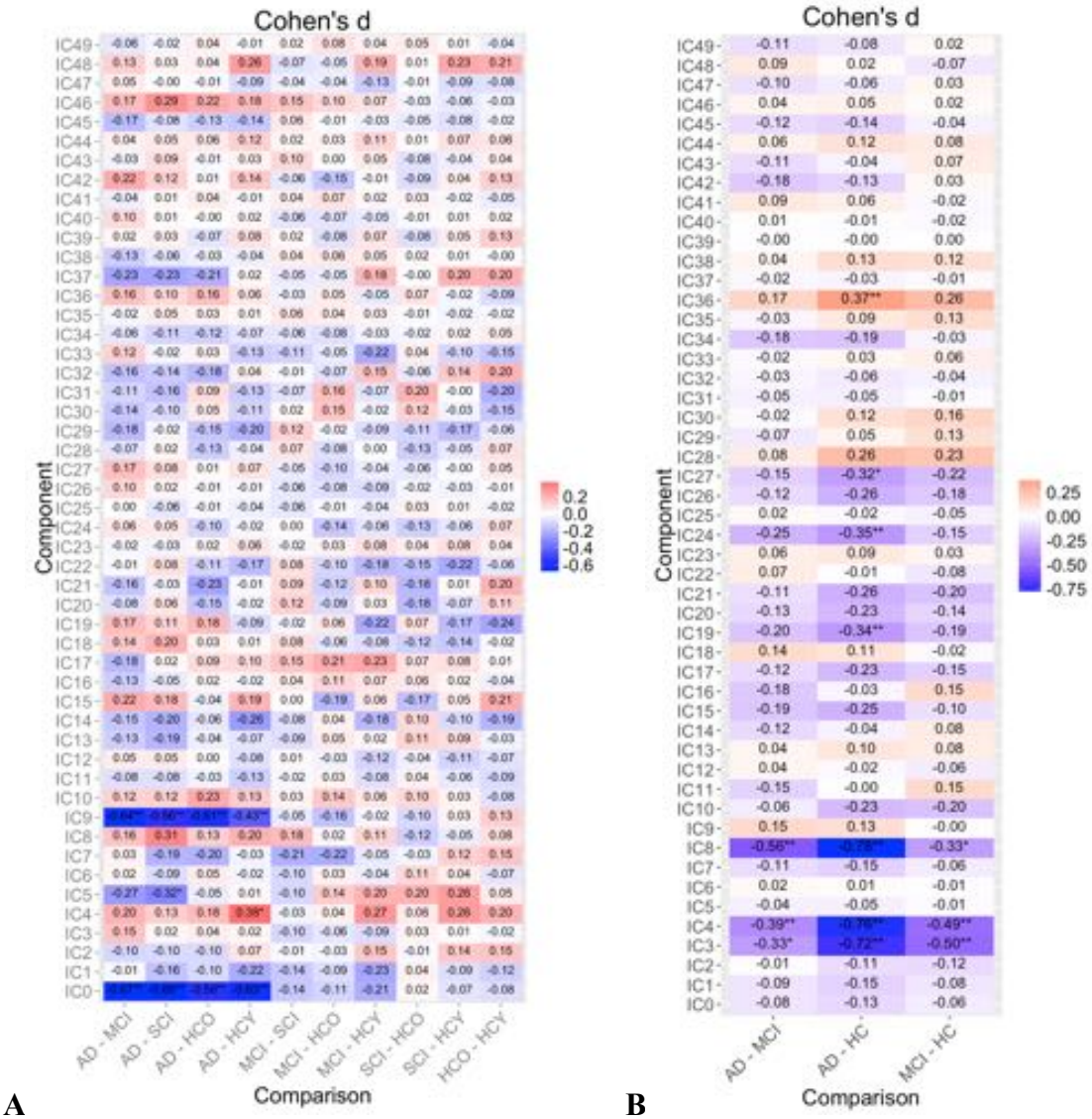


Figure S2: Effect size (Cohen's *d*) of group pairwise comparisons in (A) discovery and (B) replication samples. *, ** indicate significance levels (* $0.001<p<0.05$, ** $p<0.001$) after correcting for multiple comparison using permutation testing.

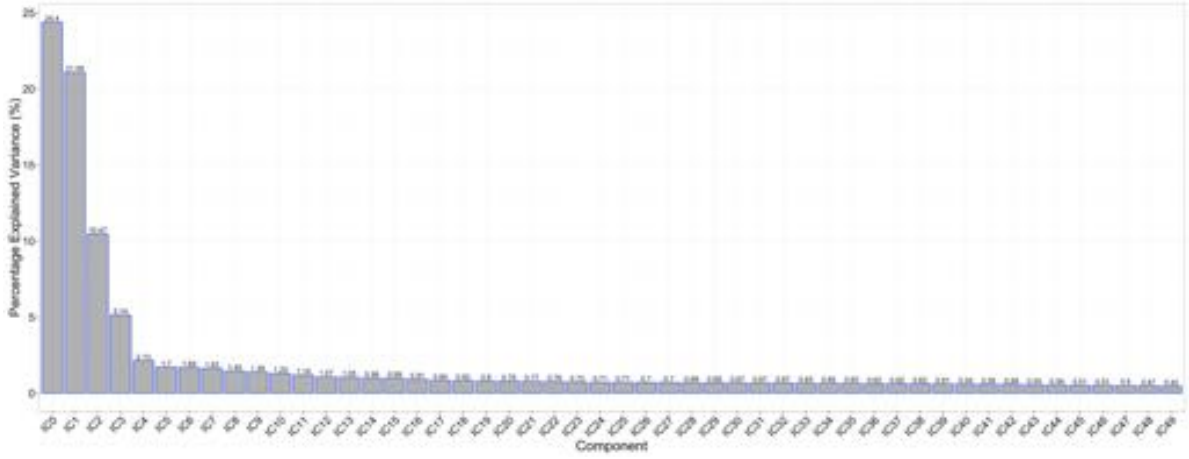


Figure S3: [Discovery sample] Percentage explained variance of all LICA components.

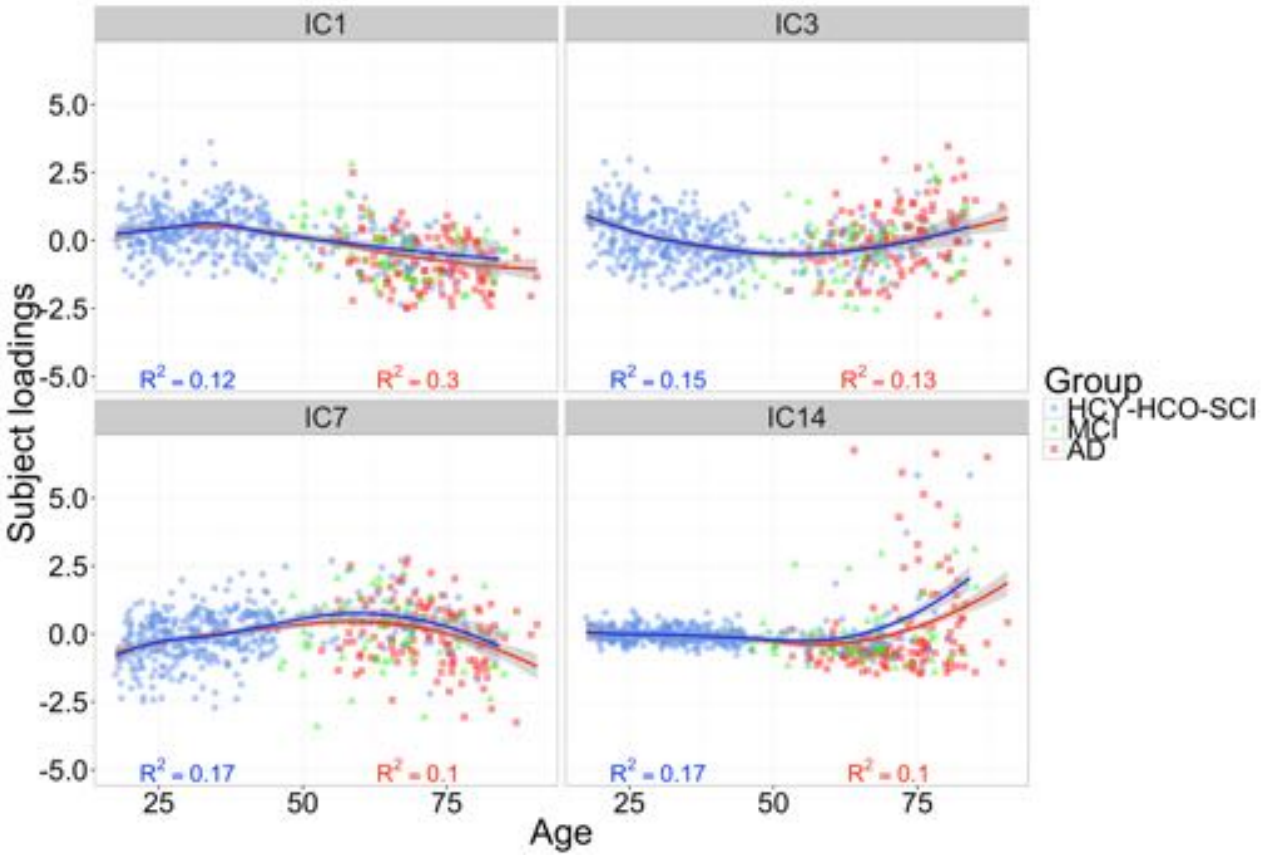


Figure S4: [Discovery sample] Scatter plot of subject loadings *versus* age for IC1,3,7,14. Only components where the LOESS fits across all datasets showed $R^2 \geq 0.1$ are shown. Blue curves represent the LOESS fit on the combined set of HCY, HCO and SCI subjects. Red curves represent the LOESS fit using all datasets.

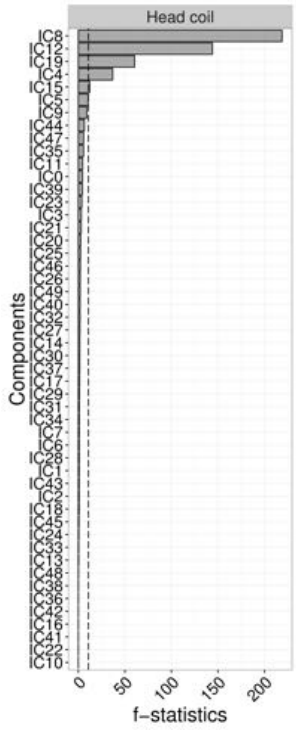
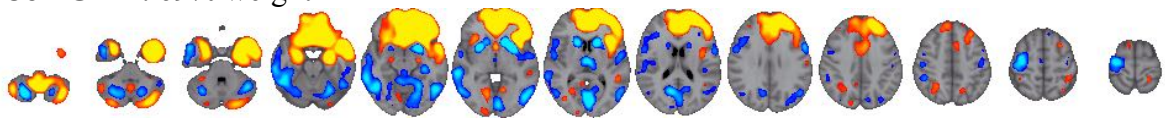
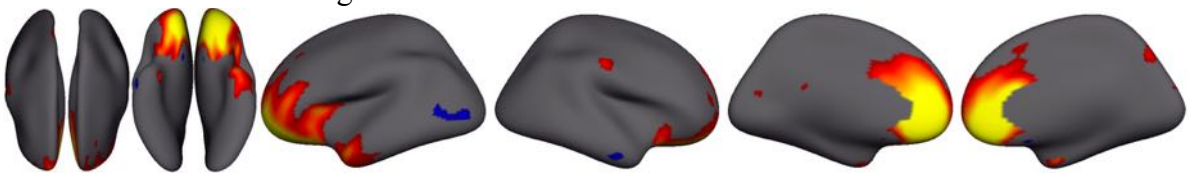


Figure S5: [Discovery sample] Main effect of head coil on the components' subject loadings. The dashed line represent the permutation-based significance threshold corrected across all components.

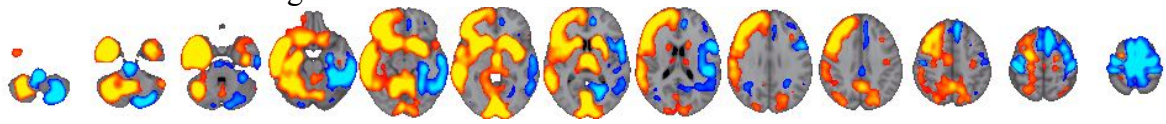
IC8 - GMD: 69% weight



IC8 - Thickness: 26% weight



IC12 - GMD: 89% weight



IC19 - GMD: 95% weight

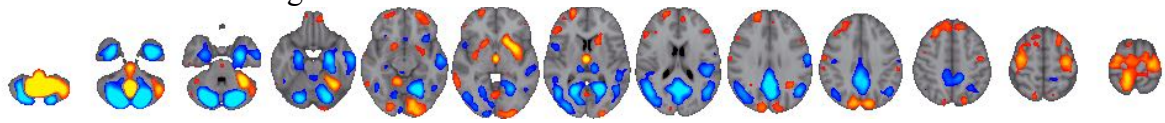
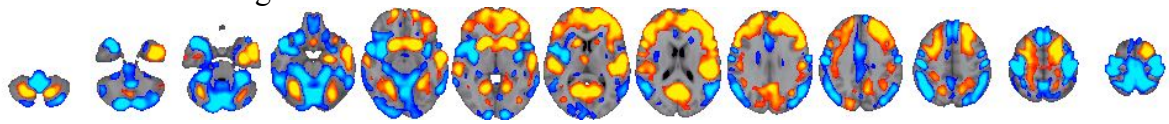
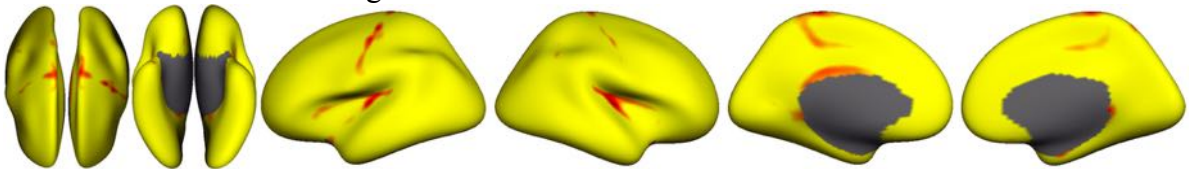


Figure S6: [Discovery sample] Spatial maps of the components capturing strong head coil effects (IC8,12,19). The spatial maps represent the thresholded z-scores ($3 < |z| < 10$). The weights indicate the relative contribution of each measure to the component at the group level. Spatial maps of measures showing negligible contribution (weight <5%) are not presented.

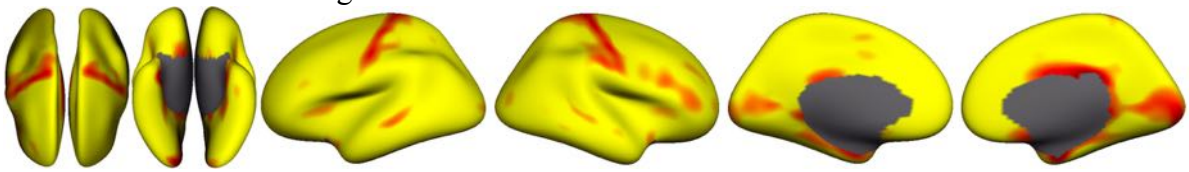
IC1 - GMD: 5% weight



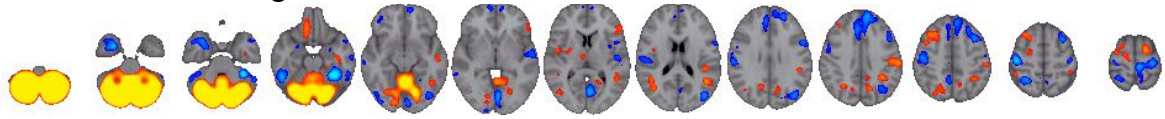
IC1 - Surface area: 93% weight



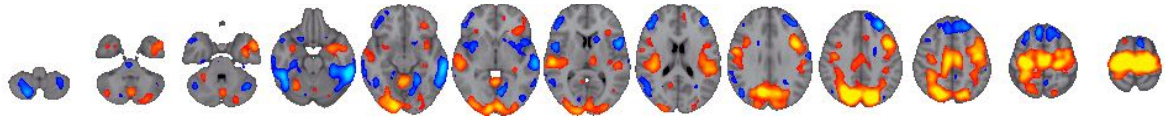
IC3 - Thickness: 96% weight



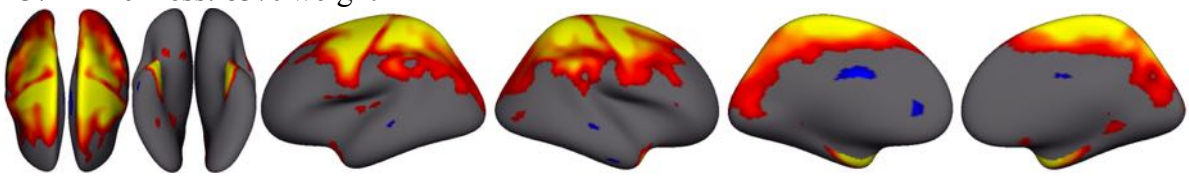
IC4 - GMD: 97% weight



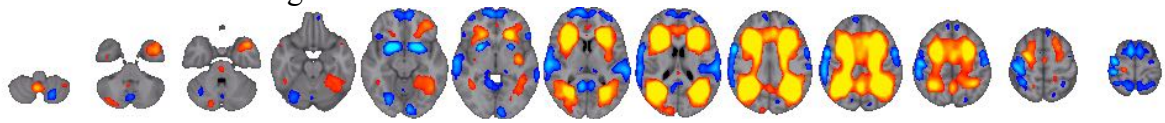
IC7 - GMD: 14% weight



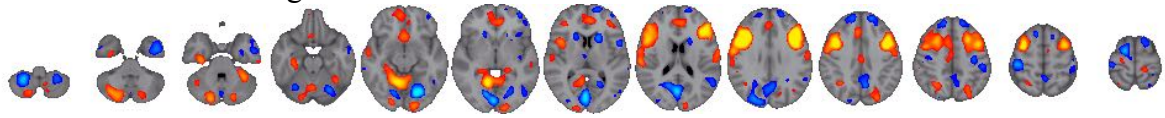
IC7 - Thickness: 83% weight



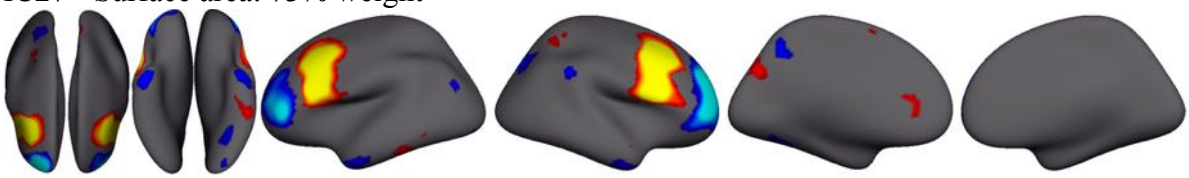
IC14 - GMD: 97% weight



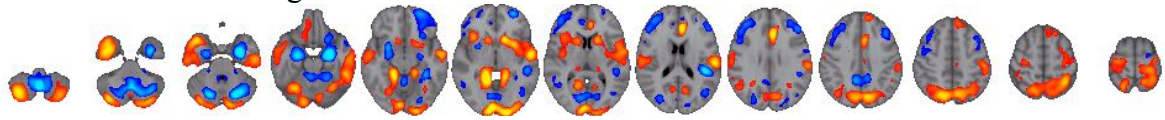
IC27 - GMD: 17% weight



IC27 - Surface area: 73% weight

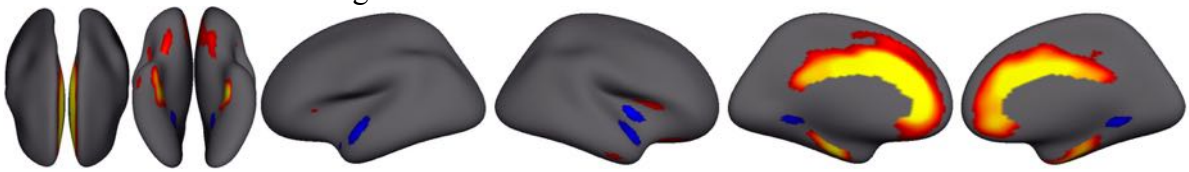


IC28 - GMD: 24% weight

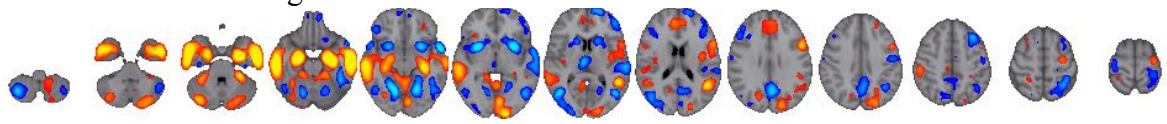


IC28 - Surface area: 13% weight

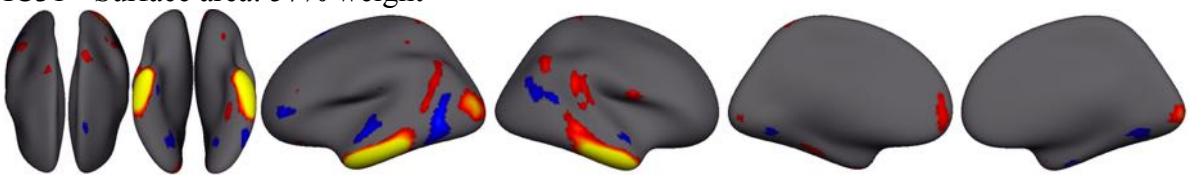
IC28 - Thickness: 58% weight



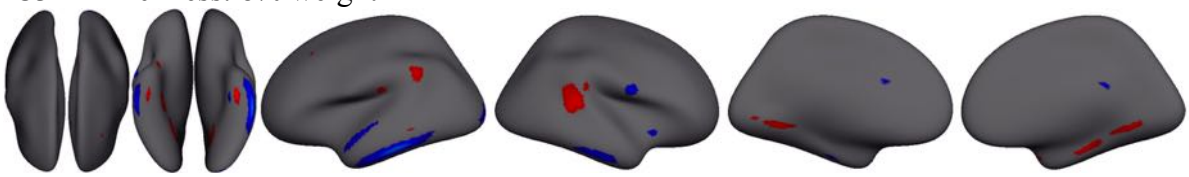
IC31 - GMD: 29% weight



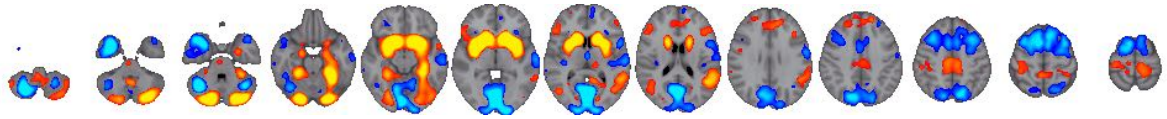
IC31 - Surface area: 57% weight



IC31 - Thickness: 8% weight

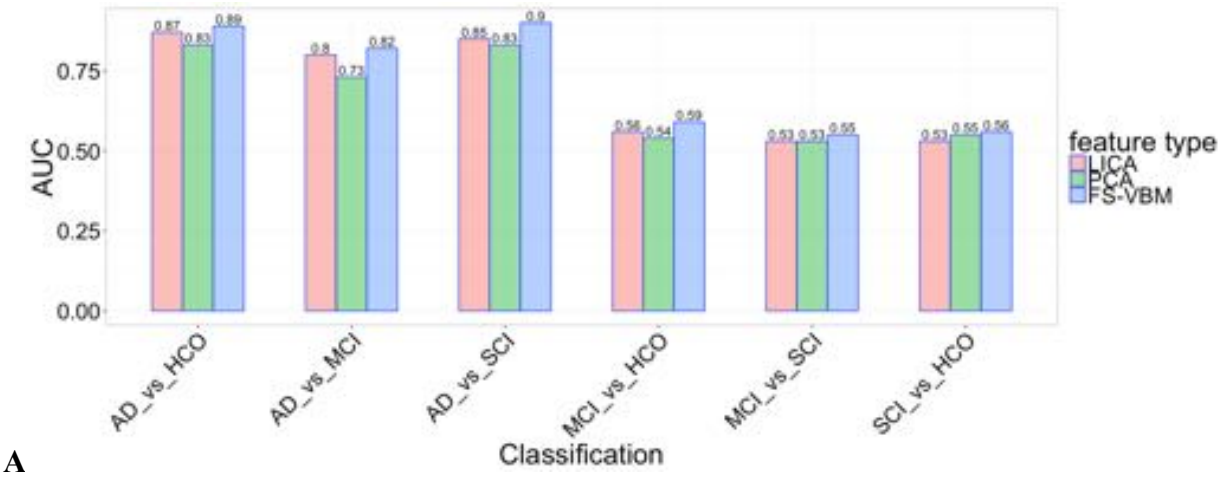


IC37 - GMD: 86% weight

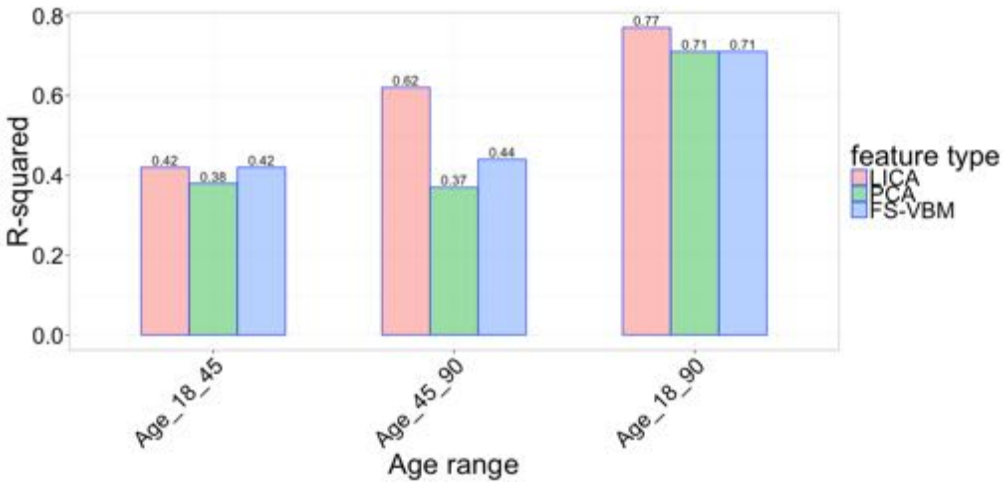


IC37 - Surface area: 7% weight

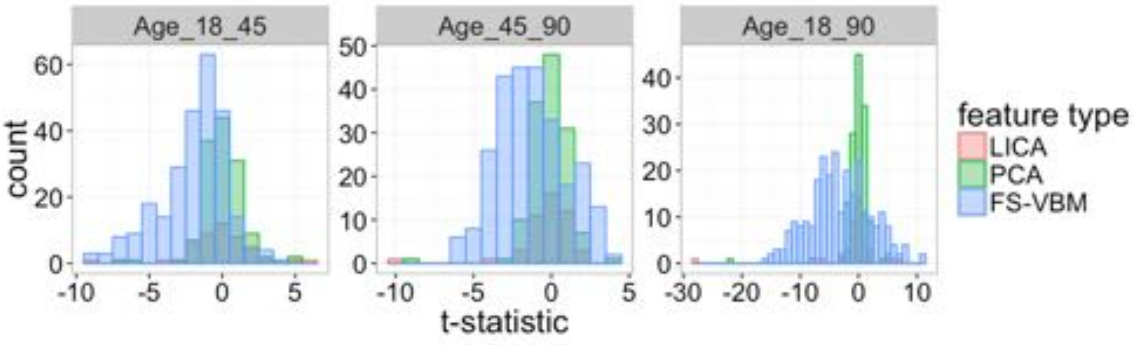
Figure S7: [Discovery sample] Spatial maps of the components rather than IC0, IC5 and IC9 sensitive to AD and age. The spatial maps represent the thresholded z-scores ($3 < |z| < 10$). The weights indicate the relative contribution of each measure to the component at the group level. Spatial maps of measures showing negligible contribution (weight <5%) are not presented.



A



B



C

Figure S8: [Discovery sample] (A) Classification performance, (B) Age prediction performance, (C) Histogram of the *t*-statistics of the age effect on either the LICA feature (*n*=47), FS-VBM feature (*n*=262) or PCA feature (*n*=136) sets.

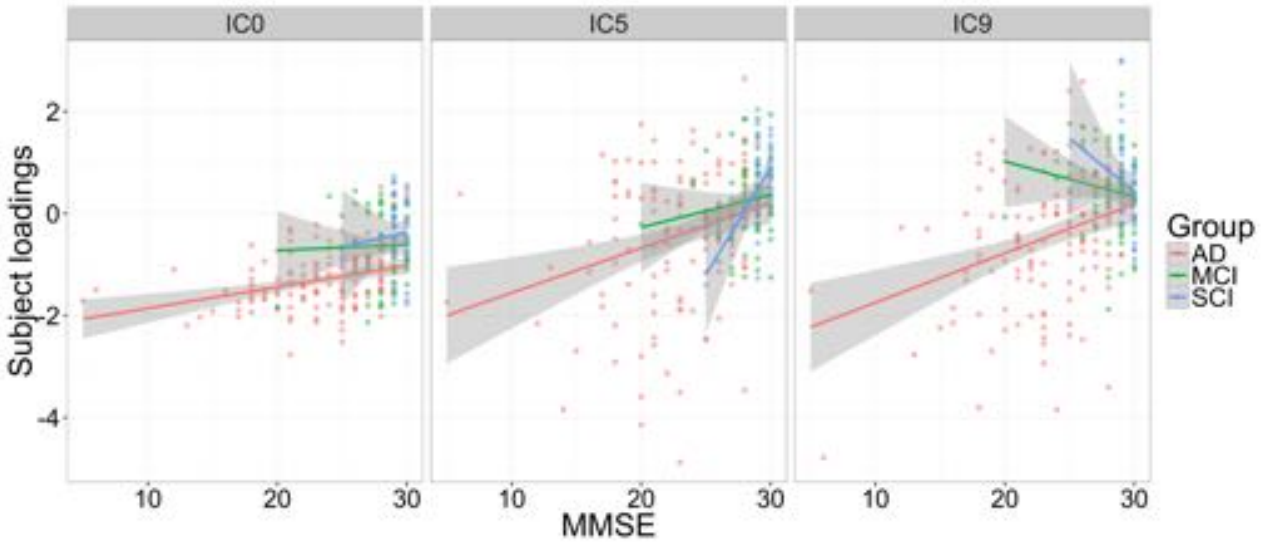


Figure S9: [Discovery sample] GLM fit within each group of subject loadings against Mini-Mental State Examination score (MMSE).

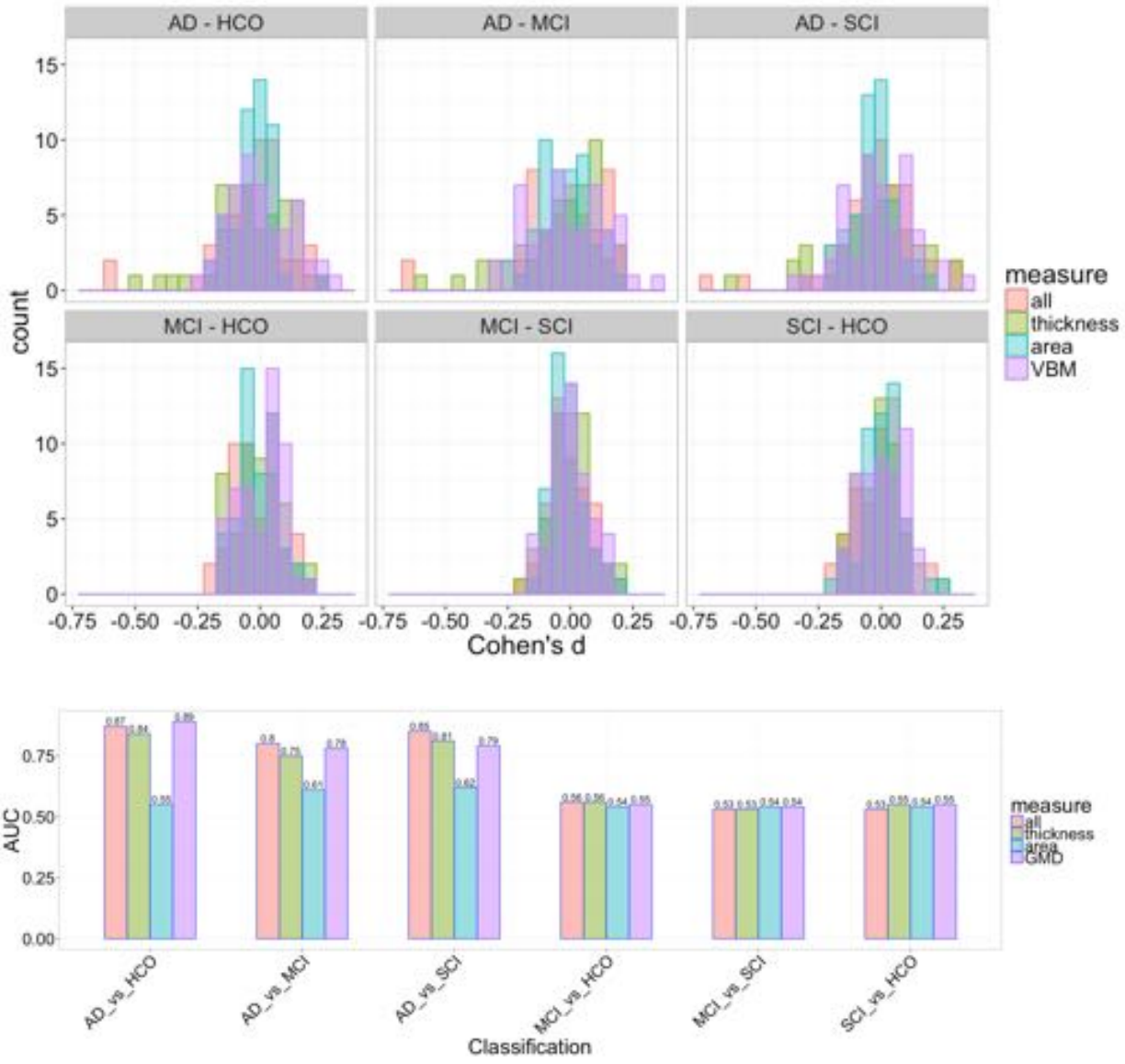
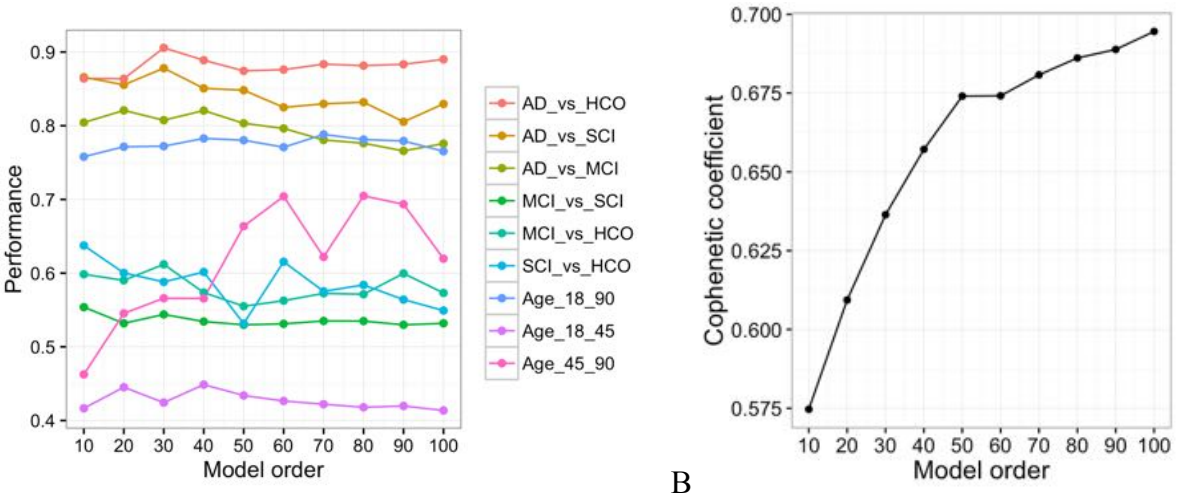


Figure S10: [Discovery sample] (A) Histogram of the Cohen's *d* and (B) Classification performance using either LICA or different ICA feature sets.



A B

Figure S11: [Discovery sample] (A) Performance (area under ROC curve for classification and R^2 between predicted and true value for age prediction) plotted as a function of LICA model order, (B) Cophenetic correlation coefficient as a function of model order.