Multi-modal Latent Factor Exploration of Atrophy, Cognitive and Tau Heterogeneity in Typical Late-Onset Alzheimer's Disease

Nanbo Sun[1], Elizabeth C Mormino[2], Jianzhong Chen[1], Mert R Sabuncu[3,4], BT Thomas Yeo[1,4,5,6], for the Alzheimer's Disease Neuroimaging Initiative[*]

[1]Department of Electrical and Computer Engineering, ASTAR-NUS Clinical Imaging Research Centre, Singapore Institute for Neurotechnology and Memory Networks Program, National University of Singapore, Singapore

[2]School of Medicine, Stanford University, Stanford, CA, USA

[3]School of Electrical and Computer Engineering, Nancy E. and Peter C. Meinig School of Biomedical Engineering, Cornell University, Ithaca, NY, USA

[4]Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown, MA, USA

[5]Center for Cognitive Neuroscience, Duke-NUS Graduate Medical School, Singapore,

[6]NUS Graduate School for Integrative Sciences and Engineering, National University of Singapore, Singapore`

Correspondence to: BT Thomas Yeo,
ECE, ASTAR-NUS CIRC, SINAPSE & MNP
National University of Singapore,
E-mail: thomas.yeo@nus.edu.sg

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

Abstract

Individuals with Alzheimer's disease (AD) exhibit significant heterogeneity across clinical symptoms, atrophy patterns, and spatial distribution of Tau deposition. Most previous studies exploring AD heterogeneity have focused on atypical clinical phenotypes, defined subtypes with a single modality, or restricted their analyses to a priori brain regions and cognitive tests. Here, we considered a data-driven Bayesian model to identify latent factors from atrophy patterns and cognitive deficits *simultaneously*, thus exploiting the rich dimensionality within each modality, and applied these methods to typical late-onset AD. Departing from most previous AD subtyping studies, our model allowed each factor to be expressed to varying degrees within an individual. This approach resulted in three atrophy-cognitive factors. The first factor was associated with atrophy of the medial temporal lobe (MTL), episodic memory deficits and disorientation to time and place ("MTL-Memory"). The second factor was associated with atrophy of the lateral temporal cortex and the default network, as well as language deficits ("Lateral Temporal-Language"). The third factor was associated with diffuse atrophy in parietal and posterior lateral temporal cortex, and visuospatial executive function deficits ("Posterior Cortical-Executive"). While the MTL-Memory and Posterior Cortical-Executive factors have been previously discussed in the literature, the Lateral Temporal-Language factor is novel and emerged only by considering atrophy and cognition jointly. The factors were replicated in an independent ADNI dementia cohort and were present in mild cognitive impaired (MCI) participants, thus capturing changes that occur prior to the onset of clinical dementia. The factors were also longitudinally stable within an individual, suggesting that these factors capture heterogeneity across patients, rather than longitudinal disease progression. Finally, using ^{18}F-AV1451 PET scans of MCI participants, we found that the atrophy-cognitive factor loadings were associated with inter-individual regional variation in Tau uptake. Furthermore, the spatial patterns of Tau uptake were consistent with the patterns of atrophy. Taken together, these results suggest that distinct atrophy-cognitive patterns exist in "typical" late-onset Alzheimer's disease and are associated with distinct patterns of Tau depositions before clinical dementia emerges.

Keywords: Alzheimer's disease, Brain atrophy, Executive function, Language, Tau imaging

1 Introduction

Alzheimer's disease (AD) is a devasting neurodegenerative disease and the most common cause of dementia. Although AD commonly presents as an amnestic syndrome (Scheltens *et al.*, 2016*b*), heterogeneous atypical AD variants have been found in many studies. For example, there is heterogeneity in atypical clinical symptoms that include dysfunction in visuospatial and language (Lam *et al.*, 2013; Ossenkoppele *et al.*, 2015), atrophy patterns (Whitwell *et al.*, 2012; Zhang *et al.*, 2016; Ferreira *et al.*, 2017) and the spatial distribution of neurofibrillary tangles (NFTs; Murray et al., 2011; Ossenkoppele et al., 2016). Several comorbid pathologies, such as Lewy body disease, transactive response DNA binding protein 43 kDa (TDP-43), vascular disease and argyrophilic grain disease (AGD), are often present in conjunction with AD pathology and may influence disease expression and progression in patients (Schneider *et al.*, 2007; Franklin *et al.*, 2015). Furthermore, there could be distinct networks of vulnerability that vary across individuals that contribute to this heterogeneity within AD.

The significant heterogeneity has motivated the exploration of AD subtypes (Murray *et al.*, 2011; Lam *et al.*, 2013; Zhang *et al.*, 2016; Young *et al.*, 2018). In most previous studies, AD subtypes were defined with a single modality, and then interrogated by comparing these subtypes with other modalities. For example, subtypes based on the spatial distribution of Tau were defined (medial temporal lobe versus cortical patterns), and their clinical symptoms and atrophy patterns were further investigated (Murray *et al.*, 2011; Whitwell *et al.*, 2012). As another example, we previously utilized a Bayesian model to identify three latent factors in typical late-onset AD dementia patients (Zhang *et al.*, 2016). Each factor was associated with a distinct pattern of brain atrophy and expressed to different degrees within each individual. Posthoc analyses suggest that the medial temporal atrophy pattern was most strongly associated with memory deficits, whereas the cortical atrophy pattern was most strongly associated with executive function deficits. Given these previously reported associations across different modalities, an approach that leverages the rich dimensionality across modalities may provide a more thorough quantitative measurement and provide novel insights of disease heterogeneity in AD.

While many of the above studies (Murray *et al.*, 2011; Lam *et al.*, 2013; Lehmann *et al.*, 2013; Ossenkoppele *et al.*, 2016) focused on heterogeneity in "atypical" AD variants (e.g.,

early-onset AD, posterior cortical atrophy, logopenic aphasia, etc), there also exists heterogeneity among typical late-onset AD dementia participants (Meda *et al.*, 2012; Noh *et al.*, 2014; Dong *et al.*, 2016; Zhang *et al.*, 2016; Ferreira *et al.*, 2017). Here, we applied an extension of our previous model (Zhang *et al.*, 2016) to investigate the heterogeneity of atrophy patterns and cognitive deficits simultaneously, by using structural MRI and neuropsychological testing scores of patients with typical late-onset AD dementia from the ADNI. The resulting multi-modal Bayesian model (Figure 1) seeks to identify latent factors associated with distinct patterns of atrophy and cognitive deficits. The advantage of combining both modalities is that the associations between atrophy patterns and cognitive deficits are automatically encouraged by the model. Thus, this approach enabled a more thorough investigation of atrophy and cognitive profiles, rather than restricting analyses to previously defined composite measures of cognition (Zhang *et al.*, 2016).

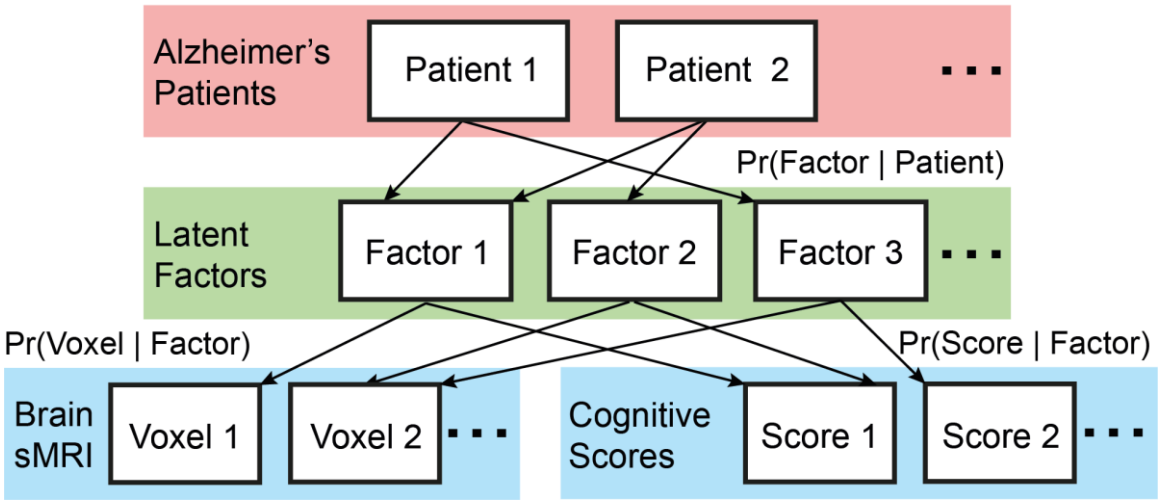

Figure 1. A Bayesian model of AD dementia patients, latent factors, brain structural MRI (sMRI) and cognitive scores. Our model allows each patient to express one or more latent factors. The factors are associated with distinct (but possibly overlapping) patterns of brain atrophy, as well as distinct (but possibly overlapping) profiles of cognitive deficits. The estimated model parameters are the probability that a patient expresses a particular factor (i.e., Pr(Factor | Patient)), the probability that a factor is associated with atrophy at a voxel (i.e., Pr(Voxel | Factor)) and the probability that a factor is associated with a cognitive score (i.e., Pr(Score | Factor)).

Consistent with our previous work (Zhang *et al.*, 2016), a key feature of our modeling approach is that it allows the possibility that multiple latent factors are expressed to varying

degrees within an individual. This is motivated by an extensive literature suggesting multiple etiologies underlying disease heterogeneity that are not mutually exclusive (Schneider *et al.*, 2009; Dickerson and Wolk, 2011; Murray *et al.*, 2011; Coutu *et al.*, 2016; Ossenkoppele *et al.*, 2016). For example, the atrophy-cognitive profiles of a patient might be 80% owing to factor 1, 15% to factor 2, and 5% to factor 3, whereas the atrophy-cognitive profiles of another patient might be 30% owing to factor 1, 65% to factor 2, and 5% to factor 3.

Although many studies have investigated disease heterogeneity at advanced stages (Murray *et al.*, 2011; Whitwell *et al.*, 2012; Byun *et al.*, 2015; Scheltens *et al.*, 2016*a*), typical late-onset AD has a long asymptomatic and prodromal stage spanning several decades before overt dementia (Scheltens *et al.*, 2016*b*). Thus, we investigated the association between atrophy patterns and cognitive deficits in MCI participants with positive β-Amyloid (Aβ+) biomarker. Furthermore, the advent of ^{18}F-AV1451, a PET tracer with high affinity for paired helical filament Tau, provides a recent opportunity to investigate regional NFTs *in vivo* (Chien *et al.*, 2013), which has been shown to more closely associate with clinical symptoms compared to patterns of Amyloid deposition (Brier *et al.*, 2016; Johnson *et al.*, 2016; Ossenkoppele *et al.*, 2016) Therefore, we further examined the association between atrophy-cognitive factor loadings with the pattern of Tau uptake among MCI participants.

There are three significant contributions in our study. In line with our previous work (Zhang *et al.*, 2016), our modeling approach allowed each participant to express multiple factors, rather than be assigned to a single subtype. Second, our approach harnessed the rich multidimensional information of all grey matter (GM) voxels and neuropsychological testing scores. For example, while we have previously uncovered a factor associated with medial temporal atrophy and memory loss (Zhang *et al.*, 2016), the current analysis revealed a factor that further implicated disorientation in time and space in addition to memory, suggesting the presence of broader hippocampal-related functional deficits in patients with AD. In addition, our approach also revealed a new factor associating lateral temporal atrophy with language deficits. Third, we show that elevations in Tau PET within the regions identified in our latent atrophy maps were associated with distinct atrophy-cognitive profiles in MCI participants, providing insights into the early expression of heterogeneity in AD.

2 Materials and methods

2.1 Overview

There are three stages of analyses (Figure 2). First, a hierarchical multimodal Bayesian model (Figure 1) was applied to typical late-onset AD dementia participants from the ADNI-GO/2 (ADNI-GO and ADNI2) database to extract factors based on patterns of atrophy and cognitive performance. The relationships between factor loadings and patient characteristics (e.g., age) were then examined. Second, we tested the generalizability, replicability and stability of the estimated factors, involving 10-fold cross-validation within ADNI-GO/2, independent replication using ADNI-1 and longitudinal data in ADNI-GO/2. Finally, we applied our model to extract factors earlier in the disease, among ADNI-GO/2 Aβ+ MCI participants. The association between factor loadings and the spatial pattern of Tau was also investigated in ADNI-2/3 MCI participants with available Tau PET.

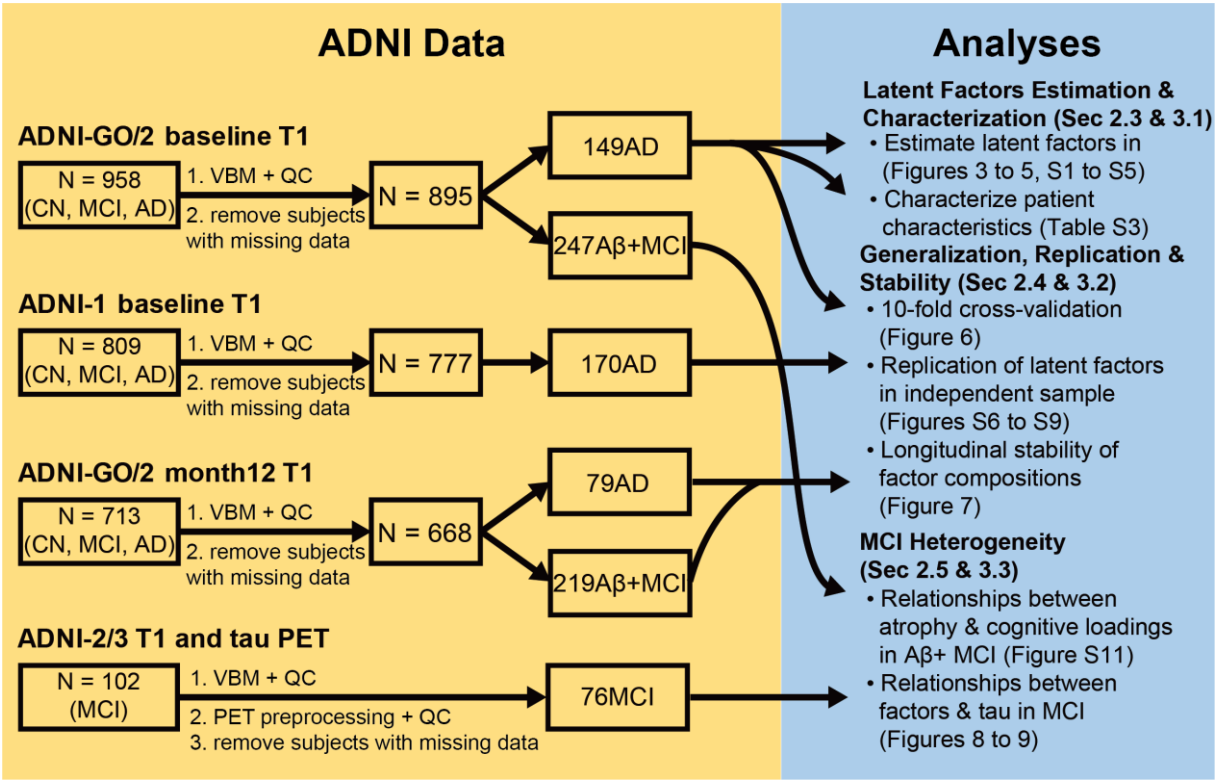

Figure 2. Overview of data and analyses. The yellow panel shows data from different ADNI cohorts and modalities. The blue panel shows the three stages of analyses in this study. Arrows indicate the data used in each analysis. We note that if Aβ+ was not specifically

mentioned for a particular cohort, then this implied that the participants were not filtered based on Aβ+. The main reason for not filtering based on Aβ+ was to maximize sample size.

2.2 Data overview

This study utilized data from the ADNI database (adni.loni.usc.edu), which was established as a public-private partnership, led by Dr. Michael W. Weiner. ADNI's main aim has been to test whether serial MRI, PET, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD. For up-to-date information, see www.adni-info.org. Written consent was obtained from all participants, and the study was approved by the Institutional Review Board at each participating institution. Demographic and clinical characteristics are presented in Table 1.

	ADNI-GO/2 baseline			ADNI-1 baseline		
	AD	MCI	CN	AD	MCI	CN
N	149	454	292	170	385	222
Age at MRI acquisition	74.6 (8.2)	71.7 (7.5)	73.0 (6.0)	75.5 (7.3)	74.9 (7.4)	76.0 (5.1)
Sex (% male)	58.4	54.2	46.2	50.6	64.2	51.4
Education (years)	15.8 (2.7)	16.1 (2.7)	16.6 (2.5)	14.7 (3.1)	15.7 (2.9)	16.0 (2.9)
% Aβ+	88.4 [N=146]	55.5 [N=445]	33.5 [N=284]	48.2	37.4	18.5
% APOE ε4 carriers	67.6 [N=145]	47.6 [N=450]	29.6 [N=291]	64.1	53.5	26.6
MMSE	23.1 (2.1)	28.1 (1.7)	29.0 (1.2)	23.5 (2.0)	27.0 (1.8)	29.1 (1.0)
CDR	0.8 (0.3)	0.5 (0.0)	0.0 (0.0)	0.7 (0.3)	0.5 (0.0)	0.0 (0.0)
	ADNI-GO/2 month 12			ADNI-2/3 Tau PET		
	AD	MCI	CN	AD	MCI	CN
N	114	353	201	_	76	_
Age at MRI acquisition	75.2 (7.7)	73.1 (7.3)	73.7 (6.2)		76.4 (7.4)	
Sex (% male)	60.5	56.4	46.8		56.6	

Education (years)	15.9 (2.6)	16.2 (2.7)	16.6 (2.5)		16.2 (2.9) [N=64]	
% Aβ+	–	–	–		55.3 [N=38]	
% APOE ε4 carriers	75.4	44.2	29.4		32.8 [N=64]	
MMSE	22.7 (3.8)	27.8 (2.0)	28.8 (1.4)		27.9 (2.1)	
CDR	0.9 (0.4) [N=113]	0.5 (0.1) [N=349]	0.1 (0.2) [N=196]		0.4 (0.2)	

Table 1. Demographic and clinical characteristics of the various ADNI cohorts utilized in this study. Abbreviations: AD=Alzheimer's disease, MCI=Mild cognitive impairment, CN=Cognitive normal, Aβ+=Amyloid beta positive, APOE=Apolipoprotein E, MMSE=Mini-mental state examination, CDR=Clinical dementia rating. Values represent mean (SD). The square brackets indicate the number of participants with available tests. For example, [N=146] indicates that there were only 146 (out of a total of 149) ADNI-GO/2 AD participants with Amyloid beta data available.

2.3 Latent Factors Estimation & Characterization

2.3.1 Structural MRI

Baseline MRI data from 158 AD dementia patients, 489 MCI participants, and 311 cognitively normal (CN) participants from ADNI-GO/2 were used to define a study specific template for voxel-based morphometry (VBM; Figure 2). The CN group was used for adjusting atrophy maps and cognitive scores before the hierarchical multimodal Bayesian model was applied to the dementia group (see below). All structural MRI (T1-weighted, 3.0T) were acquired with Sagittal MP-RAGE/IR-SPGR sequence and followed a standardized preprocessing protocol: Gradwarp, B1 non-uniformity and N3 correction (http://adni.loni.usc.edu/methods/mri-analysis/mri-pre-processing/).

The structural MRI were analyzed using the CAT12 VBM software (http://www.neuro.uni-jena.de/cat/), which is an extension of SPM12 (Penny *et al.*, 2011). Briefly, the T1 images were first normalized to a template space and segmented into GM, white matter and cerebrospinal fluid. Second, the gray and white matter images were used to create a study-specific template. Third, the template was transformed to the Montreal Neurological Institute (MNI) standard space. Fourth, the T1 images were renormalized to the study-specific template (in MNI space) and re-segmented into GM density maps. The resulting GM density images were downsampled from 1.5mm to 2mm and smoothed with a Gaussian kernel of 10-mm FWHM. The 10-mm FWHM smoothing was chosen to be

consistent with our previous work (Zhang *et al.*, 2016) and was within the typical range of 8mm to 12mm FWHM smoothing used in VBM. The downsampling reduced computational time, but did not substantially affect the resulting factors. Using the CAT12 default 8-mm FWHM smoothing (http://www.neuro.uni-jena.de/cat12/CAT12-Manual.pdf) without downsampling yielded very similar factors: mean correlation between factors with 8-mm (without downsampling) and 10-mm smoothing (with downsampling) was r = 0.95.

We applied $\log_{10}$ transformation to the GM density images and regressed out effects of age, sex and intracranial volume (ICV) from all participants. Regression coefficients were estimated from only CN participants to retain any interaction between AD and participants' characteristics (e.g., interaction between AD and age).

2.3.2 Cognitive scores

Cognitive scores from the Alzheimer's Disease Assessment Scale-Cognitive (ADAS-Cog13), Mini Mental State Exam (MMSE), Boston Naming Test, animal category of Category Fluency Test, Clock Drawing Test, Clock Copying Test, Logical Memory, Rey Auditory Verbal Learning Test and Trail Making Test were considered. See https://adni.loni.usc.edu/wp-content/uploads/2008/07/adni2-procedures-manual.pdf for detailed descriptions of these tests. In the case of MMSE, subsets of the 30 item scores were summed together into 8 domains (Table S1; Folstein *et al.*, 1975): "MMSE: Orientation to time", "MMSE: Orientation to place", "MMSE: Immediate Recall", "MMSE: Attention", "MMSE: Delayed Recall", "MMSE: Language", "MMSE: Repetition" and "MMSE: Complex Commands". This resulted in a total of 32 scores (Table S1). Some of the 32 scores exhibited limited range or low variability across participants. Therefore, we further removed some of these scores or combined them together, resulting in a total of 28 scores (Table S2). Details of this procedure are found in Supplemental Methods S1.

Among the 158 ADNI-GO/2 AD dementia participants, 9 were excluded because more than 10 cognitive scores were missing. Of the remaining 149 participants, 124 had all scores, while 25 participants had less than five missing scores. The general linear model (GLM) was used to impute the missing scores in the 25 participants. Briefly, for each participant with missing scores, we fitted a GLM to all participants (CN, MCI and AD) without missing scores, where the observed scores were independent variables and the

missing scores were dependent variables. The estimated regression coefficients were used to fill in the missing scores of the participant with missing scores.

Consistent with the structural MRI processing, we regressed out effects of age, sex and ICV from all scores using regression coefficients estimated from the CN participants.

2.3.3 Multimodal Bayesian model

We have previously utilized a hierarchical Bayesian model, latent Dirichlet allocation (LDA; Blei *et al.*, 2003), to encode the premise that each AD patient expresses one or more latent factors, associated with distinct patterns of brain atrophy (Zhang *et al.*, 2016). LDA (and its variants) have also been successfully used to extract overlapping brain networks from functional MRI (Yeo *et al.*, 2014) and meta-analytic data (Bertolero *et al.*, 2015; Yeo *et al.*, 2015). In the current analysis, we considered an extension of the LDA model to incorporate two modalities (gray matter atrophy and neuropsychological testing scores), herein referred to as multi-modality LDA (MMLDA). In this framework, each AD patient expresses one or more latent factors, each of which is associated with distinct (but possibly overlapping) atrophy patterns and distinct (but possibly overlapping) cognitive deficits (Figure 1). The model is mathematically equivalent to that proposed by Putthividhya and colleagues (2007). Details are found in Supplementary Methods S2 and S3.

We considered VBM and cognitive scores of all ADNI-GO/2 participants with AD dementia (N = 149). Since most participants were Aβ+ (88%; Table 1), we decided not to exclude the small number of Aβ- participants, in order to maximize the number of participants. The VBM and cognitive scores of 149 ADNI-GO/2 participants with AD dementia were first z-normalized with respect to the CN participants (see Supplementary Methods S4 for details). Given the normalized VBM and cognitive scores of the 149 participants, as well as a predefined number of latent factors K, a variational expectation-maximization (VEM) algorithm was used to estimate the probability of a patient expressing a latent factor [Pr(Factor | Patient)], probability that a factor was associated with atrophy at a voxel [Pr(Voxel | Factor)] and the probability that a factor was associated with a cognitive score [Pr(Score | Factor)].

An important model parameter is the number of latent factors K. We considered K = 2, 3 or 4. As will be explained in the Results, the four-factor solution was not interpretable,

and so we did not explore beyond four factors. For the remainder of this paper, we will largely focus on the three-factor solution.

2.3.4 Visualization

The probability of a voxel being associated with a factor (Pr(Voxel | Factor)) can be thought of as a brain map. For visualization, the brain map is projected from MNI152 to fsaverage space (Wu *et al.*, 2018). The fsaverage space is the official FreeSurfer surface coordinate system obtained by averaging the cortical surfaces of 40 participants (Dale *et al.*, 1999; Fischl *et al.*, 1999*a*). Subcortical structures are illustrated using coronal slices. Multiple coronal slices are also shown in supplemental figures.

2.3.5 Relationship between factors & patient characteristics

We explored how patient characteristics (age at time of structural MRI scan, education, Amyloid, and APOE genotype) varied across the three latent factors using GLMs for continuous variables and logistic regression for binary variables. In all cases, patient characteristics were the independent variables, while factor compositions were the dependent variables. Statistical tests were then performed to determine if there were overall differences across all factors, and whether there were differences between pairs of factors. See Supplementary Methods for details.

2.4 Generalization, Replication & Stability

2.4.1 Inference of atrophy loading, cognitive loading or factor loading

After the probabilistic atrophy maps Pr(Voxel | Factor) and probabilistic cognitive deficits Pr(Score | Factor) were estimated, they could be used to infer the factor composition (or loading) of a new participant using only atrophy [“atrophy loading” or Pr(Factor | Atrophy of Participant)], only cognitive scores [“cognitive loading” or Pr(Factor | Cognition of Participant)] or both atrophy and cognitive scores [“factor loading” or Pr(Factor | Participant)].

2.4.2 10-fold cross-validation

To test whether the association between cognitive profiles and atrophy patterns can generalize to new participants, we performed 10-fold cross-validation among the 149 ADNI-GO/2 AD dementia participants. Briefly, this involves randomly splitting the 149 participants into ten folds (groups) of roughly the same size. For each test fold, the multimodal Bayesian model parameters were estimated from the remaining nine training folds, and then used to infer factor loadings in the test fold participants using only atrophy ("atrophy loading") or only cognitive scores ("cognitive loading"). The prediction quality is measured by the correlation between atrophy and cognitive loadings across the test fold participants.

2.4.3 Replication in ADNI-1

The multimodal Bayesian model was also applied to an independent cohort of 170 ADNI-1 patients with AD dementia to see if the resulting factors were similar across ADNI cohorts. The model parameters were estimated completely independently from the previous analyses involving ADNI-GO/2 participants.

2.4.4 Longitudinal stability of factor compositions

To examine factor stability during disease progression, we considered the subset of 149 AD dementia and 247 Aβ+ MCI participants (from the ADNI-GO/2 cohort) with one-year follow-up structural MRI and the full set of cognitive scores, yielding 79 dementia and 219 Aβ+ MCI participants. Factor compositions of the participants were inferred using only structural MRI, only cognitive scores or both structural MRI and cognitive scores.

2.5 MCI Heterogeneity

AD has a long prodromal stage before overt dementia. Therefore, we were interested in exploring whether the atrophy-cognitive factors estimated from the AD dementia participants were relevant in an earlier disease stage (i.e., MCI).

2.5.1 Relationships between atrophy & cognitive loadings in MCI

Of the 454 ADNI-GO/2 MCI participants, 445 had Amyloid PET data available, and 247 of 445 were Aβ+ (Landau *et al.*, 2013). To explore the associations between atrophy

patterns and cognitive deficits in MCI participants, the multimodal Bayesian model parameters estimated from the ADNI-GO/2 AD dementia participants were used to infer factor loadings in the 247 Aβ+ MCI participants using only atrophy ("atrophy loading") or only cognitive scores ("cognitive loading").

To examine the relationships between cognitive and atrophy loadings, GLM was utilized. For example, the cognitive loading of factor 1 (Pr(Factor 1 | Patient's cognition) would be the dependent variable, while the atrophy loadings would be the independent variables. Age and sex were included as nuisance variables. To be more explicit, let us denote memory loading by h, medial temporal loading by m, lateral temporal loading by l, and posterior cortical loading by c. Age x_1 and sex x_2 were included as nuisance variables, so the final GLM was $\mathrm{h} = \beta_0 + \beta_l l + \beta_c c + \beta_1 x_1 + \beta_2 x_2 + \epsilon$, where β's denote the regression coefficients, and ϵ is the residual. The medial temporal loading t was implicitly modeled because $t + l + c = 1$. The GLM was repeated with Pr(Factor 2 | Patient's cognition) and then Pr(Factor 3 | Patient's cognition) as the dependent variables. For each GLM, statistical tests were performed to determine if there were overall differences across all factors, and whether there were differences between pairs of factors. See Supplementary Methods S6 for details.

2.5.2 Relationships between factors & Tau in MCI

There were 102 participants with MCI from ADNI-2/3 (ADNI-2 and ADNI-3), who had accompanying ^{18}F-AV1451 PET scans. Participants were scanned beginning at 75 min post-injection, for 30 min (6 × 5 min frames) and each scan underwent the following preprocessing steps: Co-registered Dynamic, Averaged, Standardized Image and Voxel Size and Uniform Resolution (http://adni.loni.usc.edu/methods/pet-analysis/pre-processing/).

For ADNI-2 participants, the PET scan was not acquired at baseline, but during a follow-up visit. Since structural MRI and cognitive tests were not collected for every follow-up visit, this resulted in substantial missing data. In total, 26 participants were excluded because of missing structural MRI or more than five missing cognitive scores. Of the remaining 76 participants, 29 participants had all cognitive scores, while 47 participants had three or less missing scores. Like before, missing scores were imputed using the GLM based on all ADNI-GO/2 participants (CN, MCI and AD) without missing scores. The participants

were not filtered based on Amyloid status because including only Aβ+ participants will lead to too few participants (Table 1).

All structural MR images were processed using FreeSurfer 6.0 (https://surfer.nmr.mgh.harvard.edu/) to obtain the parcellation of the cerebellum in native T1 space (Dale *et al.*, 1999; Fischl *et al.*, 1999*a*, *b*; Fischl *et al.*, 2001; Ségonne *et al.*, 2007; Greve and Fischl, 2009). All PET images were then co-registered to the corresponding T1 images. Voxelwise standardized uptake value ratio (SUVR) images were created by dividing each voxel by the mean value in the cerebellar gray matter. Finally, the PET images were transformed to the same template space as the VBM analysis and downsampled to 2mm (which was the same resolution as the GM density images).

To investigate the relationships between the spatial heterogeneity of Tau depositions and factor loadings, each probabilistic atrophy map (Pr(Voxel | Factor)) was thresholded to obtain a mask containing the top 5% of the voxels. The SUVR signal was averaged across voxels within each mask to obtain the regional Tau loading for each factor.

To determine if the regional Tau loadings were associated with the factor loadings, the GLM was applied for each factor loading. For example, the loading of factor 1 (Pr(Factor 1 | Patient)) would be the dependent variable, while the Tau loading in each of the three atrophied regions would be the dependent variables. Age, sex and total Tau were included as nuisance variables. For each GLM, statistical tests were performed to determine if there were overall differences across all factors, and whether there were differences between pairs of factors. See Supplementary Methods for details.

Finally, we correlated the Tau loadings (at each voxel) with the atrophy-cognitive factor loadings across participants, resulting in three brain maps. For example, a higher correlation at a voxel for the Posterior Cortical-Executive factor suggests that higher Tau loading at the voxel was associated with higher Posterior Cortical-Executive loading.

2.6 Statistical Analyses

In the case of logistic regression, the likelihood ratio test was employed to test for differences between factors. In the case of GLMs, the F-test was employed. In the case of Pearson's correlations (e.g., 10-fold cross-validation), p values were computed using the Student's t distribution. All tests were two-sided. Furthermore, since multiple statistical tests

(patients' characteristics analyses, 10-fold cross-validation, associations between atrophy and cognitive loadings among Aβ+ MCI participants, and associations between Tau loadings and factor loadings among MCI participants) were performed in this study, all p values were corrected using a false discovery rate (FDR) of q = 0.05.

2.7 Code and data availability

The code used in this paper, as well as the atrophy-cognitive factors are publicly available at GITHUB_LINK. The paper utilized data from the publicly available ADNI database (http://adni.loni.usc.edu/data-samples/access-data/).

3 Results

3.1 Latent Factor Estimation & Characterization

3.1.1 Latent atrophy-cognitive factors in AD dementia

Using a multimodal Bayesian model, we estimated three latent factors from ADNI-GO/2 AD dementia participants that captured covariance between patterns of atrophy and cognitive testing scores. The three latent factors involved distinct atrophy patterns (Figure 3; Figure S1) and corresponding cognitive testing profiles (Figure 4). The first factor was associated with atrophy in the medial temporal lobe (MTL) and was associated with episodic memory tests (delayed and immediate recall measures), as well as orientation measures ("MTL-Memory"). The second factor was associated with atrophy in left lateral temporal cortex and diffuse atrophy across the default network, and was associated with multiple tests assessing language ("Lateral Temporal-Language"). Finally, the third factor was associated with atrophy in the right posterior bilateral cortices and was associated with tests of executive function and visuospatial function ("Posterior Cortical-Executive").

3.1.2 Interpreting the atrophy-cognitive factors

While two of the three factors (Figure 4) were similar to previously defined composite measures of memory and executive function in ADNI (Crane *et al.*, 2012; Gibbons *et al.*, 2012), they were not identical. For example, the MTL-Memory factor included not just episodic memory measures, but also orientation to time and place, which are also well-established functions associated with the medial temporal lobe and hippocampus (O'Keefe and Dostrovsky, 1971; Hafting *et al.*, 2005). Similarly, the Posterior Cortical-Executive factor included various executive function measures such as "MMSE: attention" and "ADAS: number cancellation", which were not included in the ADNI executive function composite score (Gibbons *et al.*, 2012).

It is worth noting that each cognitive test can be thought of as an imperfect measurement of some underlying cognitive construct. The test might be imperfect in that it is only measuring part of the cognitive construct or it might involve cognitive processes beyond the construct of interest, random noise and measurement error. Here, the data-driven fusion of multiple related cognitive tests seeks to better capture underlying cognitive constructs, compared with using individual imperfect cognitive tests.

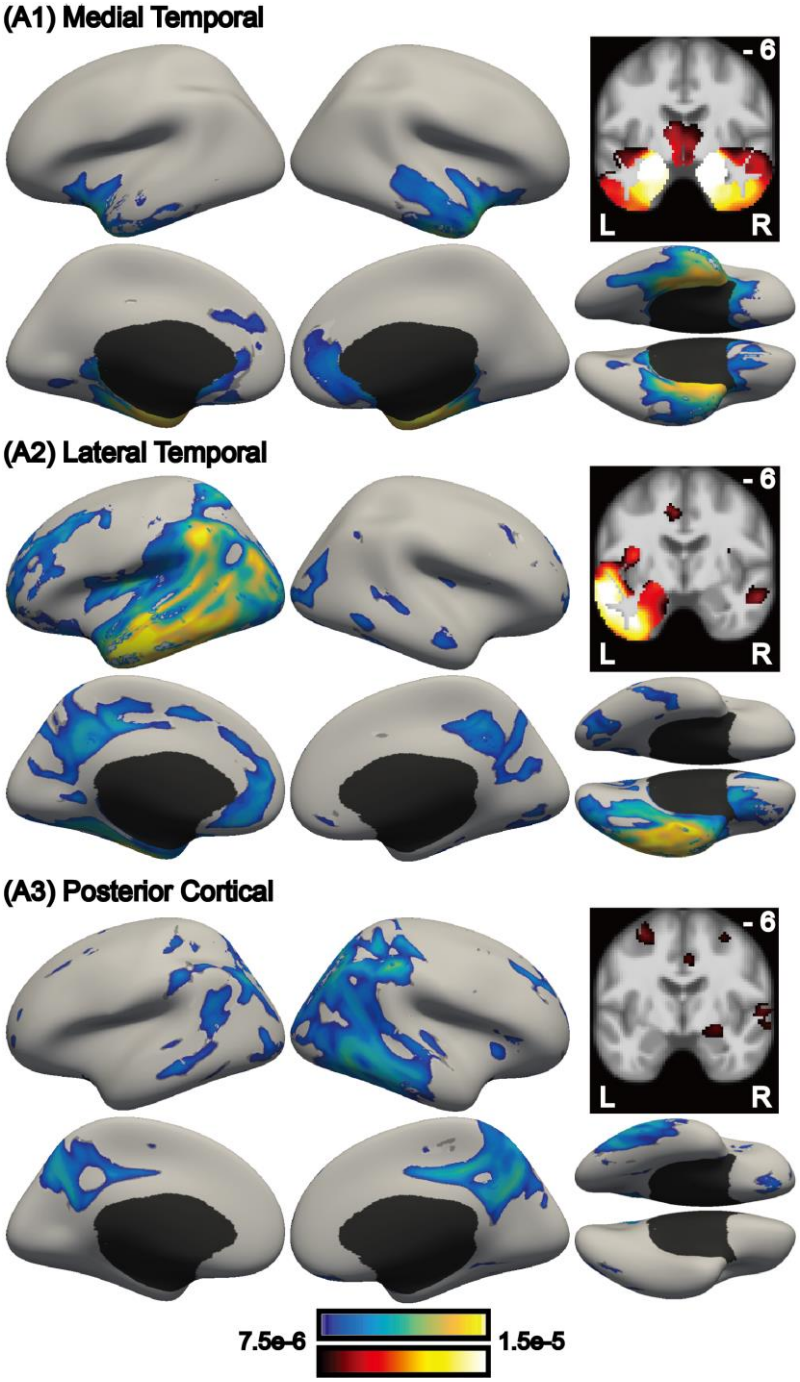

Figure 3. Probabilistic atrophy maps of three latent factors in ADNI-GO/2. Brighter color indicates higher probability of atrophy at that voxel for a particular factor (i.e., Pr(Voxel | Factor)). Each of the three factors was associated with a distinct pattern of brain atrophy. Coronal slices are found in Figure S1.

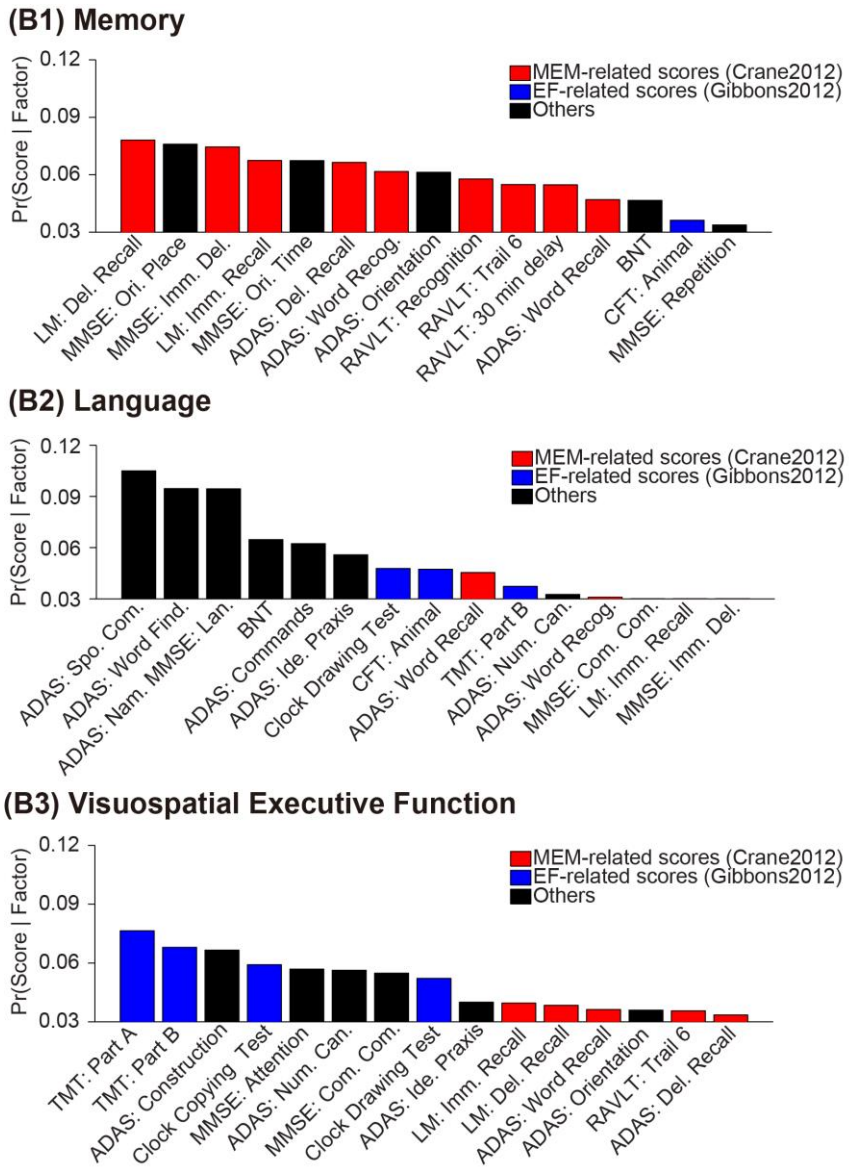

Figure 4. Probabilistic cognitive deficits of three latent factors in ADNI-GO/2. Red bars indicate cognitive scores associated with memory (as defined by Crane *et al.*, 2012). Blue bars indicate cognitive scores associated with executive function (as defined by Gibbons *et al.*, 2012). The remaining cognitive scores were colored black. Vertical axis corresponds to the probability of a cognitive score being associated with a factor (i.e., Pr(Score | Factor)). Only the top 15 scores are shown. ADAS: Alzheimer's Disease Assessment Scale; MMSE: Mini Mental State Exam; CFT: Category Fluency Test; LM: Logical Memory; RAVLT: Rey Auditory Verbal Learning Test; TMT: Trail Making Test.

It is also worth pointing out that Pr(Score | Factor) must sum to one over all scores because it is a probability distribution. Therefore, if certain cognitive tests from a particular cognitive domain were over-represented in the dataset, then on average Pr(Score | Factor) for

the particular cognitive domain would be lower. Here, there were more cognitive tests in ADNI involving memory and visuospatial executive function, so Pr(Score | Factor) was generally lower for the Posterior Cortical-Executive and MTL-Memory factors compared with the Lateral Temporal-Language factor (Figure 4).

3.1.3 Number of factors

In addition to exploring a three-factor solution, we also investigated models that defined two and four latent factors. The two-factor model revealed two factors that were very similar to the first two factors of the three-factor model. More specifically, there was one factor associated with medial temporal atrophy (Figure S2A1) and memory-related cognitive deficits (Figure S3B1), whereas the second factor was associated with posterior cortical atrophy (Figure S2A2) and visuospatial/executive function cognitive deficits (Figure S3B2). Therefore, we preferred the three-factor model to the two-factor model because of the additional Lateral Temporal – Language factor.

On the other hand, the four-factor model revealed three factors (Figures S4A1-3, S5B1-3) which were similar to factors in three-factor model, together with an additional factor with prominent medial temporal atrophy in addition to subcortical atrophy in the basal ganglia and thalamus (Figure S4A4) with a mixture of deficits across multiple cognitive domains (Figure S5B4). Because this fourth factor did not appear meaningful, by the principle of Occam's razor, we preferred the three-factor model to the four-factor model.

Because the above interpretation might be considered subjective, we also performed 10-fold cross-validation for the four-factor model (Section 3.2.1). To foreshadow the results, the fourth atrophy pattern could not successfully predict individuals' cognitive deficit profile. Thus, at least in this dataset, there were only three stable atrophy-cognitive factors, which we will focus for the remainder of this paper.

3.1.4 Characteristics of factor compositions

Examination of factor compositions revealed that the majority of the 149 AD dementia participants expressed multiple latent atrophy factors rather than predominantly expressing a single atrophy factor (Figure 5). Furthermore, there was no significant association between the factor loading and education, sex, or APOE (Table S3). However, the third factor

(Posterior Cortical-Executive) was associated with younger age (at time of scan) than the other two factors (p = 2e-4; Table S3).

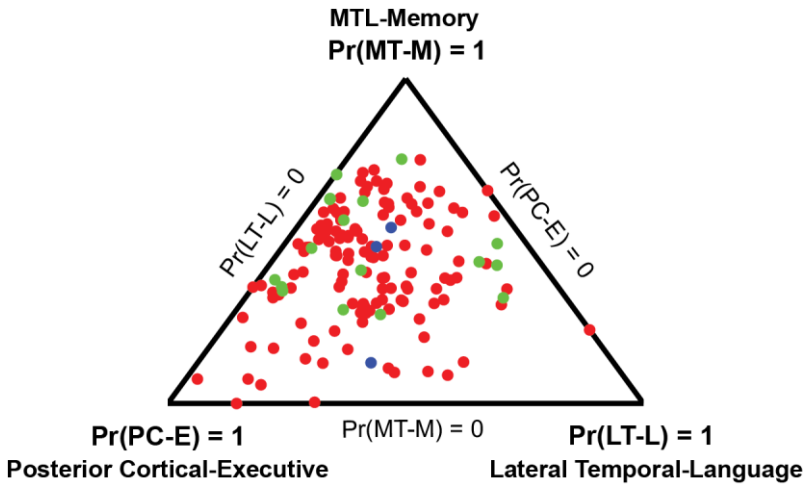

Figure 5. Factor compositions of 149 AD dementia patients. Each patient corresponds to a dot, with location (in barycentric coordinates) representing the factor composition. Color indicates Amyloid status: red for Aβ+, green for Aβ-, and blue for unknown. Corners of the triangle represent pure factors; closer distance to the respective corner indicates higher probability for the respective factor. Most dots are far from the corners, suggesting that most patients expressed multiple factors. "MT-M" indicates the MTL-Memory factor. "LT-L" indicates the Lateral Temporal-Language factor. "PC-E" indicates the Posterior Cortical-Executive factor.

3.2 Generalization, Replication & Stability

3.2.1 Atrophy pattern can be used to predict cognitive deficits

To test whether the atrophy pattern of an out-of-sample participant could be used to predict his/her cognitive deficit profile, we performed 10-fold cross-validation among the AD dementia participants in ADNI-GO/2. We found that greater medial temporal atrophy predicted worse memory function (Figure 6; r = 0.34, p = 2e-5), greater lateral temporal atrophy predicted worse language performance (Figure 6; r=0.30, p=2e-4), and greater posterior cortical atrophy predicted worse visuospatial executive function (Figure 6; r=0.30, p=3e-4).

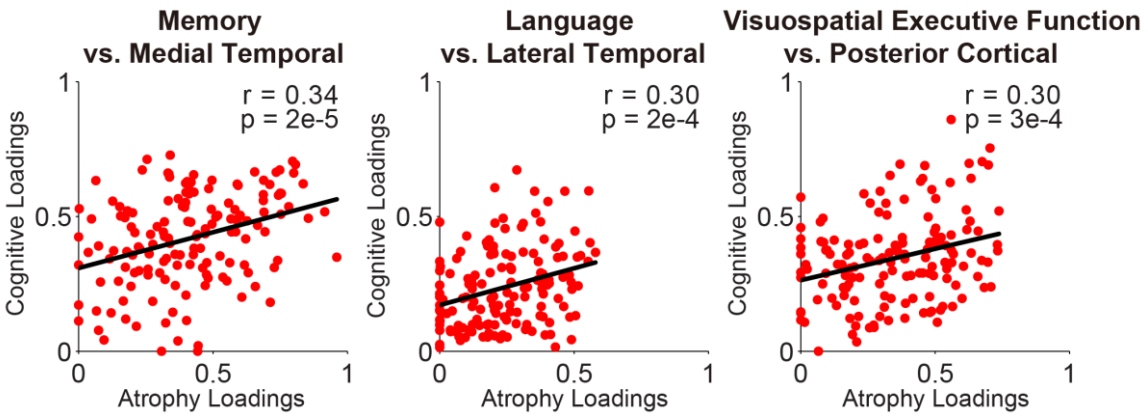

Figure 6. Atrophy pattern of a participant could be used to predict his/her cognitive deficit profile. We performed 10-fold cross-validation among the AD dementia patients in ADNI-GO/2. For each test fold, the multimodal Bayesian model parameters were estimated from the remaining nine training folds, and then used to infer factor loadings in the test fold subjects using only atrophy ("atrophy loading") or only cognitive scores ("cognitive loading"). The prediction quality is measured by the correlation of atrophy and cognitive loadings across the test fold subjects, shown above for all test folds. Each red dot represents an AD patient.

Importantly, medial temporal atrophy loading predicted memory deficit (r = 0.34) better than language (r = -0.17) or executive function (r = -0.21) deficits. Similarly, lateral temporal atrophy loading predicted predict language deficits (r = 0.30) better than memory (r = -0.17) or executive function deficits (r = -0.09). Finally, posterior cortical atrophy loading predicted visuospatial executive function deficits (r = 0.30) better than memory (r = -0.23) or language (r = -0.06) deficits.

When the 10-fold cross-validation was repeated for the four-factor solution, the first three atrophy patterns (which corresponded to the three-factor solution) successfully predicted individuals' cognitive deficit profiles with r = 0.31, 0.29 and 0.35 respectively. However, the fourth atrophy pattern failed to predict the corresponding cognitive deficit profile (r = 0.09). This further motivates our decision to focus on the three-factor model in this paper.

3.2.2 Replication in ADNI-1

We investigated whether the factors found in ADNI-GO/2 could be reproduced in an independent cohort of patients with AD from ADNI-1 (N=170). As shown in Figures S6 and S7, the spatial atrophy patterns of the three latent factors estimated in ADNI-1 were similar to ADNI-GO/2. Correlations between corresponding atrophy patterns from the two cohorts were

0.83, 0.65 and 0.63 for factors 1, 2 and 3 respectively. The main difference was more bilateral involvement of the lateral temporal lobe in ADNI-1 (Figure S6A2 compared to Figure 3A2) in the case of the lateral temporal factor.

The profiles of cognition deficits were also similar between ADNI-GO/2 and ADNI-1 (Figure S8 compared to Figure 4). For example, the top 6 cognitive scores for the MTL-Memory factor in ADNI-GO/2 was among the top 7 cognitive scores (with different ordering) for the corresponding factor in ADNI-1. Similarly, the top 5 cognitive scores for the Lateral Temporal-Language factor in ADNI-GO/2 was among the top 6 cognitive scores (with different ordering) for the corresponding factor in ADNI-1. Correlations between corresponding cognitive profiles from the two cohorts were 0.98, 0.58 and 0.83. These results suggest that although all factors were replicable, the medial temporal lobe - memory (MTL-M) factor was the most replicable, which might not be surprising given that we are focusing on AD dementia in this study.

To ensure that discrepancy in the lateralization of the lateral temporal atrophy in ADNI-1 and ADNI-GO/2 was not due to artifacts of the MMLDA algorithm, for each participant, we compute composite memory, language and visuospatial executive function scores by simply averaging the top five cognitive scores associated with each factor (Figure 4). This overall score was then correlated against voxelwise VBM data across the ADNI-1 and ADNI-GO/2 participants. The resulting correlation maps (Figure S9) were highly consistent with the atrophy patterns (Figures S1 and S7). Of particular importance is that the language composite score was correlated with atrophy in only the left temporal lobe in ADNI-GO/2, but correlated with atrophy in both left and right temporal lobes in ADNI-1. This suggests that this discrepancy was not an artifact of the MMLDA algorithm.

3.2.3 Factor stability

To determine whether factor expression remained stable over time, we examined 79 dementia and 219 Aβ+ MCI participants from ADNI-GO/2, who had both structural MRI and the full set of cognitive scores in the one-year follow-up visit. The factor compositions were highly correlated between the baseline and one-year follow up, and linear fits were close to the y = x line, suggesting that the factors were not merely reflecting disease progression (Figure 7). Interestingly, the atrophy loading (computed only using structural MRI) were

more reliable (greater Pearson's correlations) than the cognitive loading (computed only using cognitive scores).

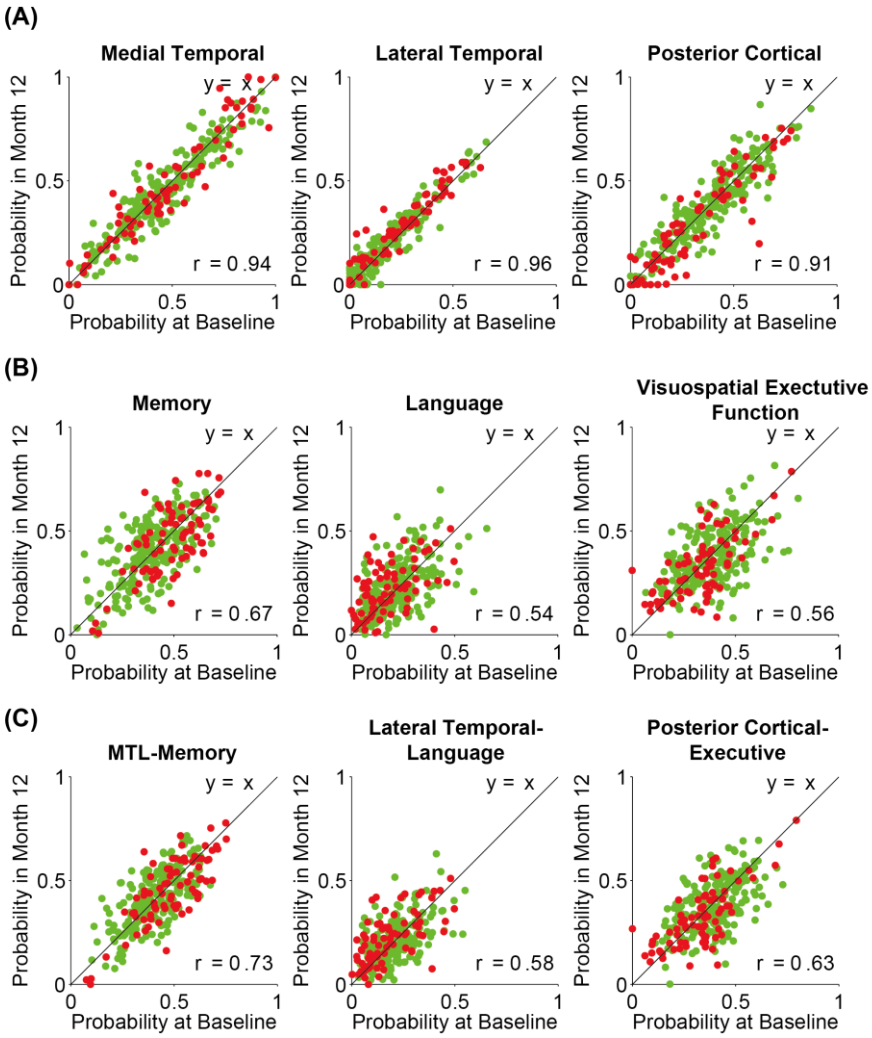

Figure 7. Stability of factor loadings after one year. Each red dot corresponds to an AD dementia participant, while each green dot corresponds to a Aβ+ MCI participant. The x axis represents the probability of each factor at baseline and the y axis represents the probability of each factor after one year. (A) Atrophy loading estimated using only structural MRI. (B) Cognitive loading estimated using cognitive scores. (C) Factor loading estimated using both structural MRI and cognitive scores. Factor compositions were highly correlated between the baseline and one-year follow up, and linear fits were close to the y = x line (black) suggesting that the factors were not merely reflecting disease stages.

3.3 MCI Heterogeneity

3.3.1 Atrophy and cognition in Aβ+ MCI

Using the model defined in the ADNI-GO/2 participants with AD dementia, we extracted factor compositions for the 247 Aβ+ participants with MCI. Like the AD dementia participants, the majority of individuals with MCI also expressed multiple factors (Figure S10).

The model parameters estimated from the AD dementia cohort were used to infer factor loading in the Aβ+ MCI participants, using only atrophy ("atrophy loading") or only cognitive scores ("cognitive loading"). GLMs were then utilized to investigate the relationship between atrophy loading and cognitive loading in the Aβ+ MCI participants (see Methods). We found that greater atrophy in posterior cortical regions (compared with lateral temporal cortex) was associated with worse visuospatial executive function ($p = 1e\text{-}3$ corrected with FDR < 0.05, Figure S11C). There was a trend that greater atrophy in posterior cortical regions (compared with MTL regions) was associated with worse visuospatial executive function ($p = 0.04$; Figure S11C). There was no significant association for memory and language deficits (Figure S11A-B).

3.3.2 Relationship between Tau deposition and factor loading

The model parameters estimated from the AD dementia cohort were used to infer factor loading in 76 MCI participants, using both atrophy and cognitive scores. GLMs were then utilized to investigate the relationship between the factor loading and the spatial distribution of Tau deposition (see Methods). Figure 8A shows that participants with higher Tau deposits in the medial temporal regions (compared with lateral temporal and posterior cortical regions) exhibited greater loading on the MTL-memory factor. On the other hand, participants with higher Tau deposits in the lateral temporal and posterior cortical regions (compared with medial temporal regions) exhibited higher loading on the Posterior Cortical-Executive factor.

Figure 9 shows the across-participant correlations between the atrophy-cognitive factor loadings and Tau deposition at each voxel. The Tau-atrophy-cognitive association maps were strikingly similar to the atrophy patterns (Figure 3). For example, higher Tau deposition in the hippocampus was associated with higher MTL-Memory factor loading (Figure 9A).

However, the correlation between the Tau-atrophy-cognitive association map (Figure 9A) and atrophy pattern (Figure 3A) was weak (r = 0.17). On the other hand, higher Tau deposition in posterior cortical regions was associated with lower Lateral Temporal-Language factor loading. Correlation between the Tau-atrophy-cognitive association map (Figure 9B) and atrophy pattern (Figure 3B) was r = -0.40. Finally, higher Tau deposition in posterior cortical regions was associated with higher Posterior Cortical-Executive factor loading. Correlation between the Tau-atrophy-cognitive association map (Figure 9C) and atrophy pattern (Figure 3C) was r = 0.42.

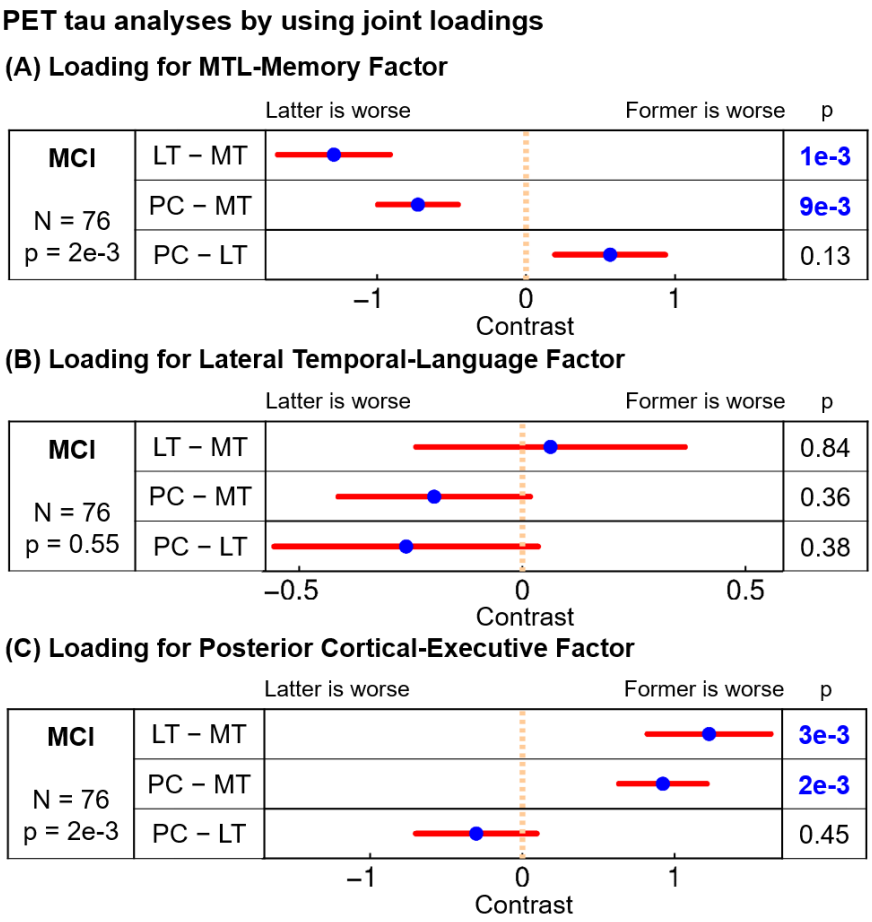

Figure 8. Associations between factor loadings and Tau deposits in atrophied regions among ADNI-2/3 MCI participants. All p values remaining significant after FDR (q < 0.05) are highlighted in blue. MT, LT, and PC indicate Tau loadings in the medial temporal lobe, lateral temporal, and posterior cortical atrophied regions respectively. Blue dots are estimated differences in regression coefficients between Tau deposits in different atrophied regions and red bars show standard errors. For example, in the bottom plot, the blue dot of "PC – MT" is on the right side (p = 2e-3), which means that participants with greater Tau deposits in the

posterior cortical regions (compared with the medial temporal regions) exhibited greater factor loading on the Posterior Cortical-Executive factor.

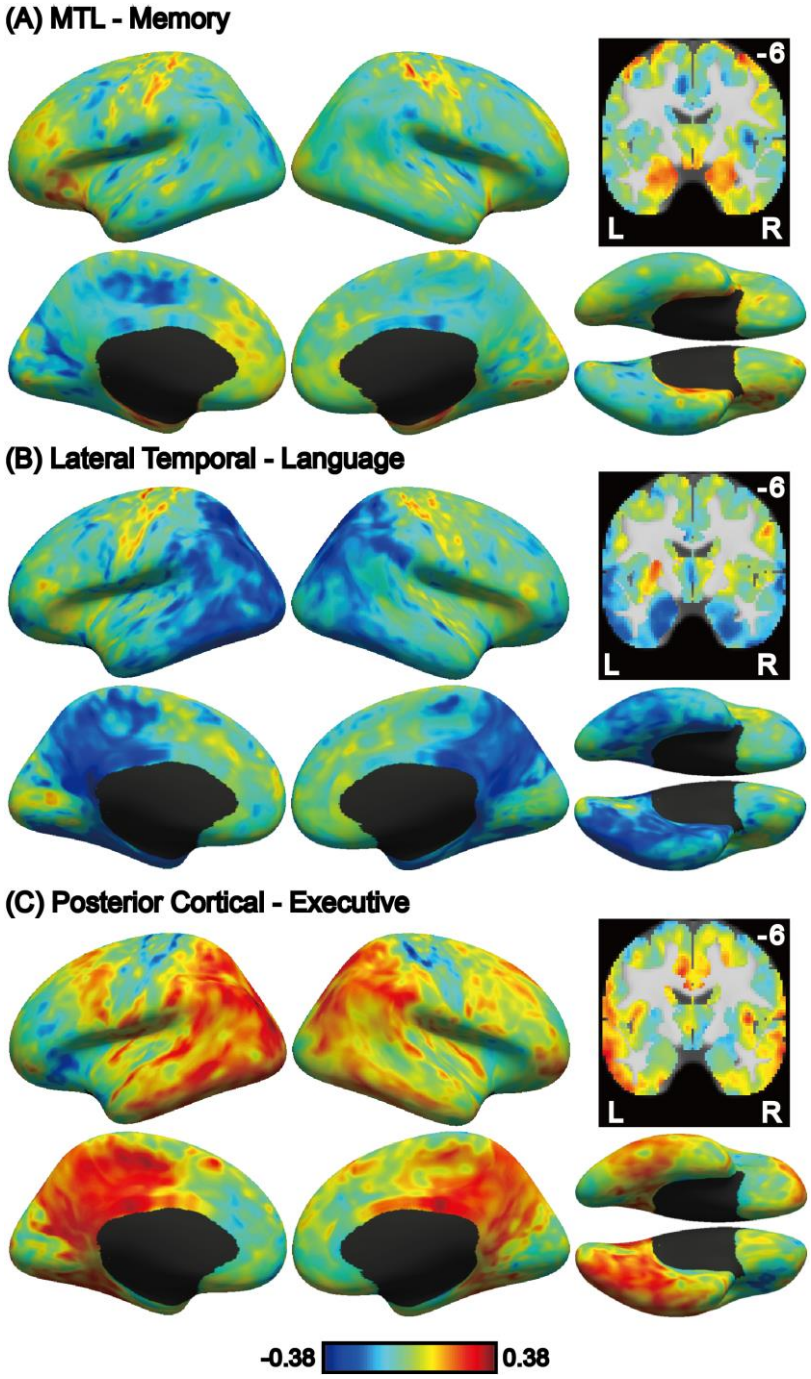

Figure 9. Correlation between atrophy-cognitive factor loading and Tau deposition (at each voxel) across participants. (A) Correlation between Tau deposition and MTL-Memory factor loading. (B) Correlation between Tau deposition and Lateral Temporal-Language factor loading. The spatial pattern is similar to the posterior cortical atrophy pattern (r = -0.40), suggesting that higher Tau deposition in posterior cortical regions is associated with lower

Lateral Temporal-Language factor loading (C) Correlation between Tau deposition and Posterior Cortical-Executive factor loading. The spatial pattern is similar to the posterior cortical atrophy pattern ($r = 0.42$), suggesting that higher Tau deposition in posterior cortical regions is associated with higher Posterior Cortical-Executive factor loading.

3.4 Comparison with CCA

Since CCA has been widely used to discover brain-behavior relationships (Smith *et al.*, 2015), we applied CCA to the ADNI-GO/2 data to investigate if meaningful atrophy-cognitive modes could be obtained (Supplementary Methods S8). Permutation testing (Smith *et al.*, 2015) did not reveal any significant CCA mode ($p > 0.13$). Nevertheless, we considered the top three components and performed 10-fold cross-validation (Supplementary Methods S8). None of the CCA mode yielded successful prediction in the out-of-sample data (Figure S12).

4 Discussion

In this study, we investigated the heterogeneity across atrophy and cognitive profiles among participants with typical late-onset Alzheimer's disease dementia using a multimodal Bayesian model. We identified three latent factors, capturing (1) medial temporal lobe atrophy that was associated with memory and orientation ("MTL-Memory"), (2) lateral temporal lobe atrophy that was associated with language ("Lateral Temporal-Language"), and (3) posterior cortical atrophy that was associated with executive function ("Posterior Cortical-Executive"). These factors were similar when re-estimated on an independent cohort of patients with AD dementia, and the associations between atrophy and cognition were also consistent in independent datasets. Among MCI patients, Tau PET signal from regions within the MTL-Memory factor were associated with loadings from the MTL-Memory factor, whereas Tau PET extracted from regions within the Posterior Cortical-Executive were associated with factor loadings from both posterior cortical and lateral temporal factors. Our results suggest that heterogeneity in AD dementia is associated with distinct patterns in atrophy and cognitive profiles, these patterns are expressed along a continuum in typical late-onset AD patients, and that these patterns are present and associated with regional Tau uptake in the stages before overt clinical dementia.

4.1 Atrophy patterns and cognitive deficits in AD

Our model revealed three factors with distinct atrophy patterns in AD patients (Figure 3). Two of these atrophy patterns were similar to results from other studies (Whitwell *et al.*, 2012; Noh *et al.*, 2014; Byun *et al.*, 2015; Zhang *et al.*, 2016; Young *et al.*, 2018), while the third (lateral temporal) atrophy pattern was a novel insight from the current modeling approach. More specifically, the first factor was associated with preferential atrophy of the medial temporal lobe; the second factor was associated with preferential atrophy of the lateral temporal cortex; and the third factor was associated with preferential atrophy of the posterior cortex. Our medial temporal atrophy pattern was similar to the "medial temporal" subtype and our posterior cortical atrophy pattern was similar to the "parietal dominant" subtype previously reported by Noh and colleagues (Noh *et al.*, 2014).

Besides the anatomical subtypes, our medial temporal atrophy pattern is similar to the "limbic predominant" Tau subtype, whereas the lateral temporal and posterior cortical atrophy

factors is similar to the "hippocampal-sparing" Tau subtype (Murray *et al.*, 2011; Whitwell *et al.*, 2012). Interestingly, the "hippocampal-sparing" subtype was associated with younger age of dementia onset (Murray *et al.*, 2011), and we found a similar younger age in our posterior cortical-executive factor (Table S3). Overall, this seems to suggest that cortical presentation is associated with younger age, whereas an MTL presentation is associated with older age. Furthermore, it is worth noting that Murray and colleagues likely examined a mixture of early-onset and late-onset AD participants. For example, the age of dementia onset for the hippocampal-sparing subtype was 63±10, suggesting a significant minority of those participants were likely early-onset AD. Thus, our results are noteworthy in that differentiation in participants' age exists even within the limited age range found in typical late-onset AD dementia.

Our previous work exploring latent atrophy factors utilizing structural MRI alone identified "Temporal", "Subcortical" and "Cortical" patterns (Zhang *et al.*, 2016). Although post-hoc analyses revealed that the temporal atrophy factor was associated with worse memory performance and the cortical atrophy factor was associated with worse executive function performance, the subcortical atrophy factor was not associated with any behavioral performance (Zhang *et al.*, 2016). Our current study expands on this work by leveraging multidimensional data across two modalities, structural MRI and cognitive scores, identifying atrophy patterns that correspond with specific cognitive profiles. For example, our medial temporal factor was associated with cognitive scores reflecting orientation, in addition to the memory tests utilized in the ADNI memory composite score (Crane et al., 2012). Similarly, our posterior cortical atrophy factor was associated with additional visuospatial executive function tests beyond the original cognitive tests utilized in the ADNI executive function composite score (Gibbon et al., 2012). Unlike our previous work, the current analysis identified a lateral temporal atrophy pattern that was associated with language scores. Interestingly, the distribution of scores across participants (Figure 5) revealed that the range of expression of the lateral temporal factor was reduced compared to the medial temporal and posterior cortical factors, suggesting that very few participants predominately express this latent pattern. This restricted expression may explain why we did not identify this pattern when exploring MRI alone (Zhang *et al.*, 2016). However, this factor emerged when considering joint correspondence between atrophy and cognition. Importantly, a previous

study (Domoto-Reilly *et al.*, 2012) has shown that among participants with AD dementia, deficits in the Boston naming test was associated with atrophy of the anterior temporal lobe, which overlapped significantly with our lateral temporal atrophy pattern. Therefore, while there are atypical AD variants involving language deficits (e.g., logopenic progressive aphasia and language variant AD; Lam *et al.*, 2013), our finding provides further evidence that varying degrees of language impairment exist among typical late-onset AD participants, and that this variation is correlated with atrophy in specific brain regions that have been previously implicated in language related impairments.

A major advantage of our current study is to flexibly examine cognitive patterns using a data-driven approach, rather than restricting associations with cognition to pre-defined factor scores. In doing so, cognitive scores were sorted based on their correspondence with atrophy patterns. The first factor was associated with multiple memory measures (Logical Memory Delayed Recall, MMSE Immediate and Delayed Recall,) and orientation (MMSE: Orientation to Place and Time), cognitive deficits that have been directly associated with the medial temporal lobe (Squire *et al.*, 2004). The second factor captured variance associated with language (ADAS: Spoken Language and Comprehension, ADAS: Word Finding, ADAS: Naming, Boston Naming Test), which have previously been associated with lateral temporal cortices (Domoto-Reilly *et al.*, 2012). Finally, the third factor captured tests reflecting visuospatial executive function (Trail Making Test: Part A, Trail Making Test: Part B, ADAS: Construction, Clock Copying Test), which have been previously shown to related to posterior parietal regions (Crutch *et al.*, 2012; Wu *et al.*, 2016). Interestingly, focal and pronounced atrophy in these factors have been consistently implicated in atypical variants of AD. Specifically, Logopenic Aphasia (LPA) is associated with atrophy in regions found within our Lateral Temporal-Language factor whereas patients with posterior cortical atrophy (PCA) show atrophy in regions within our Posterior Cortical-Executive factor (Gorno-Tempini *et al.*, 2011; Crutch *et al.*, 2012). Thus, the emergence of similar patterns in the context of typical late-onset AD in the current study suggests that vulnerability in these brain networks exists along a continuum rather than being a specific feature to atypical presentations. It is possible that atypical clinical variants represent the extreme end of a continuous distribution of deficits across different cognitive domains.

Interestingly, in addition to implicating the lateral temporal cortex, factor 2 was also associated with atrophy in multiple regions within the default network. A meta-analysis by Binder and colleagues has previously suggested that the left lateralized default network is involved in language processing and semantic memory (Binder *et al.*, 2009). Traditionally, the default network has been thought to include temporal lobe, precuneus and posterior cingulate cortex, medial prefrontal cortex, dorsal prefrontal cortex and the inferior parietal cortex (Buckner *et al.*, 2008; Yeo *et al.*, 2011). However, previous work has also suggested that the default network can be fractionated into different functional subsystems (Laird *et al.*, 2009; Andrews-Hanna *et al.*, 2010; Yeo *et al.*, 2014). Our first and second factors appeared to fractionate the default network in an interesting new way, where the medial temporal lobe was associated with the first factor, while the lateral temporal lobe and other default network regions were associated with the second factor.

Also noteworthy is that both factors 2 and 3 were associated with different portions of precuneus, which is consistent with the current understanding of the functional heterogeneity of this region (Margulies *et al.*, 2009; Fornito *et al.*, 2012). More specifically, factor 3 seems to be associated with portions of precuneus that has been previously associated with the frontoparietal control network (Figure 1 of Fornito *et al.*, 2012 and Figures 11 and 13 of Yeo *et al.*, 2011). On the other hand, factor 2 seems to be associated with the precuneus/posterior cingulate core of the default network (Figure 1 of Fornito *et al.*, 2012 and Figures 11 and 13 of Yeo *et al.*, 2011). Thus, although default network regions (in particular the precuneus) are consistently implicated in AD, the specific pattern of regions involved seems to vary across individual patients.

4.2 Association with Tau pathology

Tau aggregations into neurofibrillary tangles are a hallmark pathological feature of Alzheimer's disease (Braak and Braak, 1991), and are consistently shown to correlate more strongly with cognitive functions than Amyloid-β plaques (Nelson *et al.*, 2012; Rolstad *et al.*, 2013). Excitingly, the advent of PET ligands that bind to Tau aggregations, such as ^{18}F-AV1451 (Chien *et al.*, 2013), allows us to investigate the relationship between the spatial distribution of Tau *in vivo* throughout the course of Alzheimer's disease (James *et al.*, 2015; LaPoint *et al.*, 2017; Lowe *et al.*, 2017; Schöll *et al.*, 2017; Xia *et al.*, 2017), and has been

shown to relate to clinical symptoms associated with atypical AD such as Logopenic Aphasia and Posterior Cortical Atrophy (Ossenkoppele *et al.*, 2016).

Here, we extended the work of Ossenkoppele and colleagues by examining Tau heterogeneity in typical late-onset AD dementia, which represents a population with a more limited range of clinical deficits across non-memory domains. In the context of typical late-onset AD, associations between regional ^{18}F-AV1451 and memory (or global cognition) have been identified earlier in the "typical" AD trajectory, among participants with MCI as well as clinically normal older controls (Maass, *et al.*, 2017*a*, *b*; Vogel *et al.*, 2018). This is consistent with our findings that participants with MCI, who had higher Tau deposits in medial temporal and hippocampus regions, suffer from worse memory and orientation functions (Figure 8A). Furthermore, we extended these findings by showing that participants with MCI, who had higher Tau deposits in posterior cortical region, were associated with worse visuospatial executive function (Figure 8C). To the best of our knowledge, this is the first study showing that distinct atrophy-cognitive factors are associated with distinct patterns of Tau depositions in participants with MCI.

4.3 Subtypes and stages

Due to the cross-sectional data used in this study, it is possible that our factors reflect different disease stages along the continuum of AD progression rather than heterogeneous disease subtypes (Ritchie and Touchon, 1992). However, there are two reasons why our factors are likely to be disease subtypes, rather than simply disease stages (Fonteijn *et al.*, 2012; Young *et al.*, 2014, 2018) First, our factors reflected the impairment of distinct cognitive domains, such as memory, language and visuospatial executive function (Figure 4). Second, one-year follow-up analyses revealed that the atrophy and cognitive loadings of the three factors were stable over time within an individual (Figure 7). Thus, it is likely that these factors reflect the presence of different atrophy patterns across individuals, rather than only capturing disease progression over time.

4.4 Limitations

Our study has several limitations. First, comparison with other subtype studies were difficult due to our mixed membership modeling, while most other studies classify patients

into a single category. It would be useful to apply multiple approaches to the same dataset to enable direct comparison across methods. Second, due to limited cognitive scores and sample size, the fourth factor in K=4 did not appear biologically relevant.

5 Conclusion

By utilizing the proposed multi-modal latent Dirichlet allocation (MMLDA) model, our study revealed three latent AD factors with distinct atrophy patterns and corresponding profiles of cognitive deficits. The first factor was associated with medial temporal atrophy and deficits in memory and orientation; the second factor was associated with lateral temporal atrophy and language deficits; the third factor was associated with posterior cortical atrophy and visuospatial executive function deficits. Our approach allowed each individual to express multiple factors to various degrees, rather than assigning the individual to a single subtype. This is biologically more plausible than non-overlapping subtypes among typical late-onset AD patients, given that multiple non-mutually exclusive factors likely influence this heterogeneity (such as age, co-morbid pathologies, genetics, exposures throughout the lifespan, etc). Therefore, each participant exhibited his or her own unique factor composition, which might potentially be exploited to predict individual-specific cognitive profile that may improve disease monitoring. Finally, our study suggested that these atrophy-cognitive profiles were associated with distinct patterns of Tau deposition in mild cognitively impaired participants, highlighting the emergence of these subtypes early in the AD trajectory.

6 Acknowledgements

We are currently supported by Singapore MOE Tier 2 (MOE2014-T2-2-016), NUS Strategic Research (DPRT/944/09/14), NUS SOM Aspiration Fund (R185000271720), Singapore NMRC (CBRG/0088/2015), NUS YIA and the Singapore National Research Foundation (NRF) Fellowship (Class of 2017). Our research also utilized resources provided by the Center for Functional Neuroimaging Technologies, P41EB015896 and instruments supported by 1S10RR023401, 1S10RR019307, and 1S10RR023043 from the Athinoula A. Martinos Center for Biomedical Imaging at the Massachusetts General Hospital. Our computational work was partially performed on resources of the National Supercomputing Centre, Singapore

(https://www.nscc.sg). 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.

7 References

Andrews-Hanna JR, Reidler JS, Sepulcre J, Poulin R, Buckner RL. Functional-Anatomic Fractionation of the Brain's Default Network. Neuron 2010; 65: 550–62.

Bertolero MA, Yeo BTT, D'Esposito M. The modular and integrative functional architecture of the human brain. Proc. Natl. Acad. Sci. 2015; 112: E6798-807.

Binder JR, Desai RH, Graves WW, Conant LL. Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. Cereb. Cortex 2009; 19: 2767–96.

Blei DM, Ng AY, Jordan MI. Latent Dirichlet Allocation. J. Mach. Learn. Res. 2003; 3: 993–1022.

Braak H, Braak E. Neuropathological stageing of Alzheimer-related changes. Acta Neuropathol. 1991; 82: 239–59.

Brier MR, Gordon B, Friedrichsen K, McCarthy J, Stern A, Christensen J, et al. Tau and Aß imaging, CSF measures, and cognition in Alzheimer' s disease. Sci. Transl. Med. 2016; 8: 338ra66-.

Buckner RL, Andrews-Hanna JR, Schacter DL. The brain's default network: Anatomy, function, and relevance to disease. Ann. N. Y. Acad. Sci. 2008; 1124: 1–38.

Byun MS, Kim SE, Park J, Yi D, Choe YM, Sohn BK, et al. Heterogeneity of regional brain atrophy patterns associated with distinct progression rates in Alzheimer's disease. PLoS One 2015; 10: e0142756.

Chien DT, Bahri S, Szardenings AK, Walsh JC, Mu F, Su MY, et al. Early Clinical PET Imaging Results with the Novel PHF-Tau Radioligand [F-18]-T807. J. Alzheimer's Dis. 2013; 34: 457–68.

Coutu J-P, Goldblatt A, Rosas HD, Salat DH. White Matter Changes are Associated with Ventricular Expansion in Aging, Mild Cognitive Impairment, and Alzheimer's Disease. J. Alzheimer's Dis. 2016; 49: 329–42.

Crane PK, Carle A, Gibbons LE, Insel P, Mackin RS, Gross A, et al. Development and assessment of a composite score for memory in the Alzheimer's Disease Neuroimaging Initiative (ADNI). Brain Imaging Behav. 2012; 6: 502–16.

Crutch SJ, Lehmann M, Schott JM, Rabinovici GD, Rossor MN, Fox NC. Posterior cortical atrophy. Lancet Neurol. 2012; 11: 170–8.

Dale AM, Fischl B, Sereno MI. Cortical surface-based analysis: I. Segmentation and surface reconstruction. Neuroimage 1999; 9: 179–94.
Dickerson BC, Wolk DA. Dysexecutive versus amnesic phenotypes of very mild Alzheimer's disease are associated with distinct clinical, genetic and cortical thinning characteristics. J. Neurol. Neurosurg. Psychiatry 2011; 82: 45–51.
Domoto-Reilly K, Sapolsky D, Brickhouse M, Dickerson BC. Naming impairment in Alzheimer's disease is associated with left anterior temporal lobe atrophy. Neuroimage 2012; 63: 348–55.
Dong A, Toledo JB, Honnorat N, Doshi J, Varol E, Sotiras A, et al. Heterogeneity of neuroanatomical patterns in prodromal Alzheimer's disease: links to cognition, progression and biomarkers. Brain 2016; 140: 735–47.
Ferreira D, Verhagen C, Hernández-Cabrera JA, Cavallin L, Guo CJ, Ekman U, et al. Distinct subtypes of Alzheimer's disease based on patterns of brain atrophy: Longitudinal trajectories and clinical applications. Sci. Rep. 2017; 7: 1–13.
Fischl B, Liu A, Dale AM. Automated manifold surgery: Constructing geometrically accurate and topologically correct models of the human cerebral cortex. IEEE Trans. Med. Imaging 2001; 20: 70–80.
Fischl B, Sereno MI, Dale AM. Cortical surface-based analysis: II. Inflation, flattening, and a surface-based coordinate system. Neuroimage 1999a; 9: 195–207.
Fischl B, Sereno MI, Tootell RBH, Dale AM. High-resolution inter-subject averaging and a surface-based coordinate system. Hum. Brain Mapp. 1999b; 8: 272–84.
Folstein MF, Folstein SE, McHugh PR. 'Mini-mental state'. A practical method for grading the cognitive state of patients for the clinician. J. Psychiatr. Res. 1975; 12: 189–98.
Fonteijn HM, Modat M, Clarkson MJ, Barnes J, Lehmann M, Hobbs NZ, et al. An event-based model for disease progression and its application in familial Alzheimer's disease and Huntington's disease. Neuroimage 2012; 60: 1880–9.
Fornito A, Harrison BJ, Zalesky A, Simons JS. Competitive and cooperative dynamics of large-scale brain functional networks supporting recollection. Proc. Natl. Acad. Sci. 2012; 109: 12788–93.
Franklin EE, Perrin RJ, Vincent B, Baxter M, Morris JC, Cairns NJ. Brain collection, standardized neuropathologic assessment, and comorbidity in Alzheimer's Disease

Neuroimaging Initiative 2 participants. Alzheimer's Dement. 2015; 11: 815–22.
Gibbons LE, Carle AC, Mackin RS, Harvey D, Mukherjee S, Insel P, et al. A composite score for executive functioning, validated in Alzheimer's Disease Neuroimaging Initiative (ADNI) participants with baseline mild cognitive impairment. Brain Imaging Behav. 2012; 6: 517–27.
Gorno-Tempini ML, Hillis AE, Weintraub S, Kertesz A, Mendez M, Cappa SF, et al. Classification of primary progressive aphasia and its variants. Neurology 2011; 76: 1006–14.
Greve DN, Fischl B. Accurate and robust brain image alignment using boundary-based registration. Neuroimage 2009; 48: 63–72.
Hafting T, Fyhn M, Molden S, Moser MB, Moser EI. Microstructure of a spatial map in the entorhinal cortex. Nature 2005; 436: 801–6.
James OG, Doraiswamy PM, Borges-Neto S. PET imaging of tau pathology in Alzheimer's disease and tauopathies. Front. Neurol. 2015; 6: 1–4.
Johnson KA, Schultz A, Betensky RA, Becker JA, Sepulcre J, Rentz D, et al. Tau positron emission tomographic imaging in aging and early Alzheimer disease. Ann. Neurol. 2016; 79: 110–9.
Laird AR, Eickhoff SB, Li K, Robin DA, Glahn DC, Fox PT. Investigating the Functional Heterogeneity of the Default Mode Network Using Coordinate-Based Meta-Analytic Modeling. J. Neurosci. 2009; 29: 14496–505.
Lam B, Masellis M, Freedman M, Stuss DT, Black SE. Clinical, imaging, and pathological heterogeneity of the Alzheimer' s disease syndrome. Alzheimers. Res. Ther. 2013; 5: 1.
Landau SM, Lu M, Joshi AD, Pontecorvo M, Mintun MA, Trojanowski JQ, et al. Comparing positron emission tomography imaging and cerebrospinal fluid measurements of β-amyloid. Ann. Neurol. 2013; 74: 826–36.
LaPoint MR, Chhatwal JP, Sepulcre J, Johnson KA, Sperling RA, Schultz AP. The association between tau PET and retrospective cortical thinning in clinically normal elderly. Neuroimage 2017; 157: 612–22.
Lehmann M, Madison CM, Ghosh PM, Seeley WW, Mormino E, Greicius MD, et al. Intrinsic connectivity networks in healthy subjects explain clinical variability in Alzheimer's disease. Proc. Natl. Acad. Sci. 2013; 110: 11606–11.
Lowe VJ, Wiste HJ, Senjem ML, Weigand SD, Therneau TM, Boeve BF, et al. Widespread brain tau and its association with ageing, Braak stage and Alzheimer's dementia. Brain 2017;

141: 271–87.
Maass A, Landau S, Horng A, Lockhart SN, Rabinovici GD, Jagust WJ, et al. Comparison of multiple tau-PET measures as biomarkers in aging and Alzheimer's disease. Neuroimage 2017a; 157: 448–63.
Maass A, Lockhart SN, Harrison TM, Bell RK, Mellinger T, Swinnerton K, et al. Entorhinal tau pathology, episodic memory decline and neurodegeneration in aging. J. Neurosci. 2017b; 38: 530–43.
Margulies DS, Vincent JL, Kelly C, Lohmann G, Uddin LQ, Biswal BB, et al. Precuneus shares intrinsic functional architecture in humans and monkeys. Proc. Natl. Acad. Sci. 2009; 106: 20069–74.
Meda SA, Narayanan B, Liu J, Perrone-Bizzozero NI, Stevens MC, Calhoun VD, et al. A large scale multivariate parallel ICA method reveals novel imaging-genetic relationships for Alzheimer's disease in the ADNI cohort. Neuroimage 2012; 60: 1608–1621.
Murray ME, Graff-Radford NR, Ross OA, Petersen RC, Duara R, Dickson DW. Neuropathologically defined subtypes of Alzheimer's disease with distinct clinical characteristics: A retrospective study. Lancet Neurol. 2011; 10: 785–96.
Nelson PT, Alafuzoff I, Bigio EH, Bouras C, Braak H, Cairns NJ, et al. Correlation of Alzheimer Disease Neuropathologic Changes With Cognitive Status: A Review of the Literature. J. Neuropathol. Exp. Neurol. 2012; 71: 362–81.
Noh Y, Jeon S, Lee JM, Seo SW, Kim GH, Cho H, et al. Anatomical heterogeneity of Alzheimer disease: based on cortical thickness on MRIs. Neurology 2014; 83: 1936–44.
O'Keefe J, Dostrovsky J. The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. Brain Res. 1971; 34: 171–5.
Ossenkoppele R, Pijnenburg YAL, Perry DC, Cohn-Sheehy BI, Scheltens NME, Vogel JW, et al. The behavioural/dysexecutive variant of Alzheimer's disease: clinical, neuroimaging and pathological features. Brain 2015; 138: 2732–49.
Ossenkoppele R, Schonhaut DR, Schöll M, Lockhart SN, Ayakta N, Baker SL, et al. Tau PET patterns mirror clinical and neuroanatomical variability in Alzheimer's disease. Brain 2016; 139: 1551–67.
Penny WD, Friston KJ, Ashburner JT, Kiebel SJ, Nichols TE. Statistical parametric mapping: the analysis of functional brain images. Elsevier; 2011.

Ritchie K, Touchon J. Heterogeneity in senile dementia of the Alzheimer type: Individual differences, progressive deterioration or clinical sub-types? J. Clin. Epidemiol. 1992; 45: 1391–8.
Rolstad S, Berg AI, Bjerke M, Johansson B, Zetterberg H, Wallin A. Cerebrospinal fluid biomarkers mirror rate of cognitive decline. J. Alzheimer's Dis. 2013; 34: 949–56.
Scheltens NME, Galindo-Garre F, Pijnenburg YAL, van der Vlies AE, Smits LL, Koene T, et al. The identification of cognitive subtypes in Alzheimer's disease dementia using latent class analysis. J. Neurol. Neurosurg. Psychiatry 2016a; 87: 235–43.
Scheltens P, Blennow K, Breteler MMB, de Strooper B, Frisoni GB, Salloway S, et al. Alzheimer's disease. Lancet 2016b; 388: 505–17.
Schneider JA, Arvanitakis Z, Bang W, Bennett DA. Mixed brain pathologies account for most dementia cases in community-dwelling older persons. Neurology 2007; 69: 2197–204.
Schneider JA, Arvanitakis Z, Leurgans SE, Bennett DA. The neuropathology of probable Alzheimer disease and mild cognitive impairment. Ann. Neurol. 2009; 66: 200–8.
Schöll M, Ossenkoppele R, Strandberg O, Palmqvist S, Study SB, Jögi J, et al. Distinct 18F-AV-1451 tau PET retention patterns in early- and late-onset Alzheimer's disease. Brain 2017; 140: 2286–94.
Ségonne F, Pacheco J, Fischl B. Geometrically accurate topology-correction of cortical surfaces using nonseparating loops. IEEE Trans. Med. Imaging 2007; 26: 518–29.
Smith SM, Nichols TE, Vidaurre D, Winkler AM, J Behrens TE, Glasser MF, et al. A positive-negative mode of population covariation links brain connectivity, demographics and behavior. Nat. Neurosci. 2015; 18: 1–7.
Squire LR, Stark CEL, Clark RE. The Medial Temporal Lobe. Annu. Rev. Neurosci. 2004; 27: 279–306.
Vogel JW, Mattsson N, Iturria-Medina Y, Strandberg OT, Schöll M, Dansereau C, et al. Data-driven approaches for tau-PET imaging biomarkers in Alzheimer's disease. Hum. Brain Mapp. 2018: 1–14.
Whitwell JL, Dickson DW, Murray ME, Weigand SD, Tosakulwong N, Senjem ML, et al. Neuroimaging correlates of pathologically defined subtypes of Alzheimer's disease: A case-control study. Lancet Neurol. 2012; 11: 868–77.
Wu J, Ngo GH, Greve D, Li J, He T, Fischl B, et al. Accurate nonlinear mapping between

MNI volumetric and FreeSurfer surface coordinate systems. Hum. Brain Mapp. 2018: 1–16.
Wu Y, Wang J, Zhang Y, Zheng D, Zhang J, Rong M, et al. The Neuroanatomical Basis for Posterior Superior Parietal Lobule Control Lateralization of Visuospatial Attention. Front. Neuroanat. 2016; 10: 32.
Xia C, Makaretz SJ, Caso C, McGinnis S, Gomperts SN, Sepulcre J, et al. Association of in vivo [18F]AV-1451 tau PET imaging results with cortical atrophy and symptoms in typical and atypical Alzheimer disease. JAMA Neurol. 2017; 74: 427–36.
Yeo BTT, Krienen FM, Chee MWL, Buckner RL. Estimates of segregation and overlap of functional connectivity networks in the human cerebral cortex. Neuroimage 2014; 88: 212–27.
Yeo BTT, Krienen FM, Eickhoff SB, Yaakub SN, Fox PT, Buckner RL, et al. Functional specialization and flexibility in human association cortex. Cereb. Cortex 2015; 25: 3654–72.
Yeo BTT, Krienen FM, Sepulcre J, Sabuncu MR, Lashkari D, Hollinshead M, et al. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. J. Neurophysiol. 2011; 106: 1125–65.
Young AL, Marinescu R-V V, Oxtoby NP, Bocchetta M, Yong K, Firth N, et al. Uncovering the heterogeneity and temporal complexity of neurodegenerative diseases with Subtype and Stage Inference. Nat. Commun. 2018; 9: 4273.
Young AL, Oxtoby NP, Daga P, Cash DM, Fox NC, Ourselin S, et al. A data-driven model of biomarker changes in sporadic Alzheimer's disease. Brain 2014; 137: 2564–77.
Zhang X, Mormino EC, Sun N, Sperling RA, Sabuncu MR, Yeo BTT. Bayesian model reveals latent atrophy factors with dissociable cognitive trajectories in Alzheimer's disease. Proc. Natl. Acad. Sci. 2016; 113: E6535-44.

Multi-modal Latent Factor Exploration of Atrophy, Cognitive and Tau Heterogeneity in Typical Late-Onset Alzheimer's Disease

Supplementary Material

This supplementary material is divided into **Supplementary Methods, Supplementary Figures and Tables.**

Supplementary Methods

This section provides additional information about processing, mathematical modeling and statistical procedures. Section S1 provides mathematical details about the extension of latent dirichlet allocation (LDA) from unimodal to multimodal data. Section S2 explains how the multimodal Bayesian model parameters are estimated. Section S3 explains how structural MRI and cognitive scores can be mapped as inputs to the multimodal Bayesian model. Section S4 provides more information about the setup and statistical tests for general linear modeling (GLM) and logistic regression of patient characteristics. Section S5 provides more information about the setup and statistical tests for general linear modeling (GLM) of atrophy and cognitive loadings. Finally, section S6 provides more information about the setup and statistical tests for general linear modeling (GLM) of tau loading with atrophy-cognitive factor loading.

S1. Removing and combining low-variance cognitive scores

Table S4 shows the standard deviation of the cognitive scores across CN participants within the ADNI-1 and ADNI-Go/2 cohorts. The five cognitive scores with lowest variance were consistent across both cohorts and corresponded to "ADAS: Recall Instructions", "MMSE: Language", "ADAS: Comprehension", "MMSE: Immediate Recall" and "ADAS: Spoken Language". The following cognitive scores were added together based on our understanding that they likely tested similar cognitive functions: (1) "MMSE: Language" and "ADAS: Naming", (2) "MMSE: Immediate Recall" and "MMSE: Delayed Recall" and (3) "ADAS: Spoken Language" and "ADAS: Comprehension". We did not think that the remaining cognitive score "ADAS: Recall Instructions" was tapping into similar cognitive function as other scores, so we remove the score from consideration. The variance of the combined scores are shown in Table S5.

S2. Extension of LDA to multimodal data.

The LDA model was originally developed to automatically discover latent topics from a corpus of text documents. The model assumes that each document is an unordered collection of words associated with a subset of *K* latent topics. Each topic is represented by a probability distribution over a dictionary of words. Given a collection of documents, there exist algorithms to estimate the probability of a dictionary word given a topic Pr(Word | Topic) and the probability that a topic is associated with a particular document Pr(Topic | Document). The LDA model is useful, because it allows a document to be associated with multiple topics (which can be shared across documents) and each topic to be associated with multiple words (which can be shared across topics). In our previous work, we applied LDA to discover atrophy factors (Zhang *et al.*, 2016). To achieve this, one can think of AD patients as text documents, atrophy factors as topics, and MNI152 voxels as dictionary words. Correspondingly, each patient expresses one or more latent atrophy factors to different extents [Pr(Factor | Patient)], and each factor is associated with atrophy at multiple voxels to different extents [Pr(Voxel | Factor)].

Here, an extension the LDA model was utilized to incorporate two modalities (MRI gray matter atrophy and neuropsychological testing scores), herein referred to as multi-modality LDA (MMLDA). Different extensions of LDA have been proposed to handle multiple modalities (Barnard *et al.*, 2003; Blei and Jordan, 2003). Our extension is basically the same as that of Putthividhya and colleagues (Putthividhya *et al.*, 2007), although the following exposition has been devised to (hopefully) be more intuitive for the likely readers of this work. To this end, we discuss the extension of LDA to multimodal data by first staying within the text analysis domain.

More specifically, to extend LDA to handle multiple modalities, we can think of each text document being written in multiple languages with imperfect translations, so that the information contents are not identical across languages. We can think of each language as one modality and for simplicity, we will only assume two languages (or modality), e.g., English and Spanish. Given a corpus of text documents written in two languages (English and Mandarin) and for a fixed number of topics K, multimodal LDA seeks to estimate the probability a text document involves a topic (θ, i.e., Pr(topic | document)), the probability that a topic is associated with an English word (β^1, i.e., Pr(English word | Topic)) and the probability that a topic is associated a Spanish word (β^2, i.e., Pr(Spanish word | Topic)).

Suppose there are K topics and D documents. Then θ is a $D \times K$ matrix, where each row sums to one. θ_d is the d-th row of the θ matrix and the k-th element indicates the probability that the d-th document involves the k-th topic. Suppose there are V words in the English vocabulary. Then β^1 is a $K \times V$ matrix, where each row sums to one. β_k^1 is the k-th row of the β^1 matrix and the v-th element ($\beta_{k,v}^1$) indicates the probability that the k-th topic is associated with atrophy at the v-th English word. On the other hand, suppose there are C words in the Spanish dictionary. Then β^2 is a $K \times C$ matrix, where each row sums to one. β_k^1 is the k-th row of the β^1 matrix and the c-th element ($\beta_{k,c}^1$) indicates the probability that the k-th topic is associated with the c-th Spanish word.

The model (Fig. S13) assumed a prior (Dirichlet(α)) on the topic composition (θ_d) for each document. Given the topic composition (θ_d), the total word count (N_d^1) associated with the English translation of the d-th document, and the total word count (N_d^2) associated with the Spanish translation of the d-th document, the two translations of the d-th document are generated as follows.

1. Each word in the English translated document is generated independently. For the n-th English word
    a. Sample a topic $z_{d,n}^1$ using θ_d (d-th row of the θ matrix)
    b. Sample the word $w_{d,n}^1$ using the distribution specified in the $z_{d,n}^1$-th row of the β^1 matrix.
2. Each word in the Spanish translated document is generated independently. For the n-th Spanish word
    a. Sample a topic $z_{d,n}^2$ using θ_d (d-th row of the θ matrix)
    b. Sample the word $w_{d,n}^2$ using the distribution specified in the $z_{d,n}^2$-th row of the β^2 matrix.

The above generative process specifies the mathematical model exactly. The following section will describe how the model parameters are estimated.

S3. Estimation and inference of the multimodal Bayesian model

Given the text documents translated into both English and Spanish, and a pre-defined number of topics K, the goal is to estimate the probability that a text document involves a topic (θ, i.e., Pr(topic | document)), the probability that a topic is associated with an English word (β^1, i.e., Pr(English word | Topic)) and the probability that a topic is associated a Spanish word (β^2, i.e., Pr(Spanish word | Topic)). Like the original LDA paper (Blei *et al.*, 2003), we utilized the variational expectation-maximization (VEM) algorithm to estimate the parameters.

The algorithm involves iterating between the variational E-step and M-step until convergence. In the E-step, α, β^1, β^2 are assumed fixed and the following variational parameters are updated by iterating the following two equations until convergence.

$$\phi_{d,n,k}^{l} \propto \beta_{k,v(n_{ld})}^{l} exp\left(\psi(\gamma_{d,k}) - \psi\left(\sum_{i=1}^{K} \gamma_{d,i}\right)\right) \quad (1)$$

$$\gamma_{d,k} = \alpha + \sum_{l=1}^{2} \sum_{n=1}^{N_d^l} \phi_{d,n,k}^{l}, \ (2)$$

where ϕ is the variational parameter for z, so we can interpret $\phi_{d,n,k}^{1}$ as the posterior probability that the n-th word of the English translation of the d-th document belonged to the k-th topic. Similarly, we can interpret $\phi_{d,n,k}^{2}$ as the posterior probability that the n-th word of the Spanish translation of the d-th document belonged to the k-th topic. γ is the variational parameter for θ, so we can interpret $\gamma_{d,k}$ as the posterior probability of the d-th document exhibiting the k-th topic. $v(n_{1d})$ indexes the word (in the English dictionary) that the n-th word of the English translation of the d-th document corresponded to, while $v(n_{2d})$ indexes the word (in the Spanish dictionary) that the n-th word of the Spanish translation of the d-th document corresponded to. ψ is the digamma function (the first derivative of the log Gamma function).

In the M-step, ϕ and γ are assumed fixed. The hyperparameter α is updated using the Newton-Raphson algorithm (Blei *et al.*, 2003) and β is updated as follows

$$\beta_{k,v}^{l} \propto \sum_{d=1}^{D} \sum_{n=1}^{N_d^l} \phi_{d,n,k}^{l} \, w_{d,n}^{l,v} \quad (3)$$

where $w_{d,n}^{1,v}$ is 1 if the n-th English word in the d-th document corresponded to the v-th dictionary word and 0 otherwise. $w_{d,n}^{2,v}$ is 1 if the n-th Spanish word in the d-th document corresponded to the v-th dictionary word and 0 otherwise. Looking at Eq. (1) to Eq. (3), we observe that each document can be summarized by the number of times each English word occurs and the number of times each Spanish word occurs.

Suppose we are given a new document that has been translated into both English and Spanish. Assuming that we have already estimated α, β^1, β^2, then we can simply apply the E-step (Eq. (1) and Eq. (2)) till convergence to estimate the topic composition of the new document.

We can also infer the topic composition of a new document based on a single language (modality). This can be accomplished by applying the E-step with the following slight modification:

$$\gamma_{d,k}^{l} = \alpha + \sum_{n=1}^{N_d^l} \phi_{d,n,k}^{l}. \quad (4)$$

Comparing Eq. (2) and Eq. (4), we note that Eq. (2) summed over both languages, while Eq. (4) does not. Upon convergence, we can interpret $\gamma_{d,k}^{1}$ as the topic loading of the document based only on the English translation, and $\gamma_{d,k}^{2}$ as the topic loading of the document based only on the Spanish translation.

S4. Application of Bayesian model to structural MRI and cognitive scores

To map the multimodal LDA to VBM and cognitive scores, we can think of English dictionary words as corresponding to MNI voxels and Spanish dictionary words as correspond to the cognitive scores. Furthermore, multimodal LDA assumes that a document is summarized by the number of times each English dictionary word appears in the English translation and the number of times each Spanish dictionary word appears in the Spanish translation. Because word count is discrete and non-negative, so the log-transformed GM density images and cognitive scores have to be thresholded and discretized.

More specifically, for each voxel of the log-transformed GM density images, z transformation (using the mean of CN participants and variance of all participants) was applied.

The resulting z-scores were flipped in sign, so that a positive z-score at a particular voxel of an individual would indicate atrophy at the voxel relative to CN participants. The same z-normalization procedure was applied to cognitive scores, such that a positive cognitive z-score would indicate poorer performance than CN participants.

Finally, the atrophy and cognitive z-scores were thresholded and discretized. If the threshold was set at zero (negative values set to zero), too much statistical power would be lost in the case of the cognitive scores (Fig. S14). Therefore, we added an offset of one to all z-scores before thresholding at zero. Varying this offset (e.g., adding two instead of one) did not significantly modify the results.

Furthermore, there were hundreds of thousands of voxels, but only 28 cognitive scores. To balance the two modalities, the cognitive z-scores were multiplied by the ratio between number of voxels and number of cognitive scores. Finally, both atrophy and cognitive z scores were multiplied by 10 and rounded to the nearest integer.

To summarize, for each participant, the brain atrophy pattern is represented by a non-negative integer at each voxel and the cognitive deficit profile is represented by a non-negative integer for each cognitive test. A more positive integer at a voxel would indicate greater atrophy, while a more positive integer for a cognitive score would indicate worse cognitive performance.

To estimate the model parameters, the VEM algorithm was run with 20 different random initializations and the result with highest log likelihood was chosen as the final estimate.

S5. GLM and logistic regression for patient characteristics

We apply GLM to continuous patient characteristics (e.g., age) and logistic regression to binary patient characteristics (e.g., sex) to evaluate their relationships with factor loadings. To illustrate the GLM, let's denote age by y, factor 1 loading by f_1, factor 2 loading by f_2, and factor 3 loading by f_3. Then the GLM was $\mathrm{y} = \beta_0 + \beta_2 f_2 + \beta_3 f_3 + \epsilon$, where $\beta_0, \beta_2, \beta_3$ denote regression coefficients and ϵ is the residual. Factor 1 loading f_1 was implicitly modeled because $f_1 + f_2 + f_3 = 1$.

Statistical tests involved null hypotheses of the form $H\beta = 0$, where $\beta = [\beta_0,\ \beta_2,\ \beta_3]^T$, and H is the linear contrast. First, we performed an omnibus test of overall differences across all factor loadings with $H = [0, 1, 0; 0, 0, 1]$. We then tested for pairwise differences between factor loadings. For example, $H = [0, 1, 0]$ tested the difference between factor 2 and factor 1 loadings;

$H = [0, 0, 1, 0, 0]$ tested the difference between factor 3 and factor 1 loadings; $H = [0, -1, 1, 0, 0]$ compared factor 3 and factor 2 loadings.

To illustrate the logistic regression, let's denote sex by y, factor 1 loading by f_1, factor 2 loading by f_2, and factor 3 loading by f_3. The regression model was $\log(\mu/(1-\mu)) = \beta_0 + \beta_2 f_2 + \beta_3 f_3 + \epsilon$, where μ is the probability of female, $\beta_0, \beta_2, \beta_3$ denote regression coefficients and ϵ is the residual. Intuitively, the linear combination $\beta_0 + \beta_2 f_2 + \beta_3 f_3$ predicts the probability of female ($y = 1$); $\exp(\beta_0)$ reflects the odds ratio for factor 1, $\exp(\beta_2)$ reflects the ratio of odds ratio between factors 3 and 1, and $\exp(\beta_3)$ reflects the ratio of odds ratio between factors 2 and 1.

The likelihood ratio test was utilized to determine whether the probabilities of sex were different across factors. More specifically, the test involved comparing the likelihood of an appropriate restricted model to the original model. We first performed a statistical test of overall differences across factors. In this case, the restricted model $\log(\mu/(1-\mu)) = \beta_0 + \epsilon$ was fitted to the data, and the resulting likelihood was compared with that of the original model. We then tested for differences between the factors. For example, to compare factor 2 and factor 3 loadings, the restricted model was specified as $\log(\mu/(1-\mu)) = \beta_0 + \beta_{23}(f_2 + f_3) + \epsilon$ because $\beta_2 = \beta_3$ under the null hypothesis. To compare factor 1 and 2 loadings, the restricted model was $\log(\mu/(1-\mu)) = \beta_0 + \beta_3 f_3 + \epsilon$, because $\beta_2 = 0$ under the null hypothesis.

S6. GLM of atrophy and cognitive loadings in Aβ+ MCI participants

To evaluate the associations between atrophy patterns and cognitive deficits in MCI participants, we applied the GLM for each cognitive loading. For example, let's denote memory loading by h, medial temporal loading by m, lateral temporal loading by l, and posterior cortical loading by c. Age x_1 and sex x_2 were included as nuisance variables, so the final GLM was $\mathrm{h} = \beta_0 + \beta_l l + \beta_c c + \beta_1 x_1 + \beta_2 x_2 + \epsilon$, where β's denote the regression coefficients, and ϵ is the residual. The medial temporal loading t was implicitly modeled because $t + l + c = 1$.

Similar to the previous section, statistical tests involved null hypotheses of the form $H\beta = 0$, where $\beta = [\beta_0,\ \beta_l,\ \beta_c,\ \beta_1,\ \beta_2]^T$, and H is the linear contrast. First, we performed an test of overall differences across all atrophy factor loadings with $H = [0, 1, 0, 0, 0; 0, 0, 1, 0, 0]$. We then tested for pairwise differences between atrophy loadings. For example, $H =$

$[0, 1, 0, 0, 0]$ tested the differences between lateral temporal and medial temporal atrophy loadings; $H = [0, 0, 1, 0, 0]$ compared posterior cortical and medial temporal atrophy loadings; $H = [0, -1, 1, 0, 0]$ compared posterior cortical and lateral temporal atrophy loadings.

S7. GLM of tau loading and factor loading

To evaluate the association between tau loadings and factor loadings, we applied the GLM for each factor loading. For example, let's denote MTL-Memory loading by h, medial temporal tau loading τ_m, lateral temporal tau loading τ_l and posterior cortical tau loading τ_c. The nuisance variables included age x_1, sex x_2, and total tau x_3, which was obtained by averaging tau SUVR within the gray matter of each subject. Hence, the GLM was $h = \beta_0 + \beta_m\tau_m + \beta_l\tau_l + \beta_c\tau_c + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon$, where β's denote the regression coefficients, and ϵ is the residual. In this case, we note that $\tau_m + \tau_l + \tau_c$ is not necessarily a constant, so all three tau loadings are included in the GLM.

Statistical tests included null hypotheses of the form $H\beta = 0$, where $\beta = [\beta_0, \beta_m, \beta_l, \beta_c, \beta_1, \beta_2, \beta_3]^T$, and H is the linear contrast. First, we performed an omnibus test of overall differences across all tau loadings with $H = [0, -1, 1, 0, 0, 0, 0; 0, -1, 0, 1, 0, 0, 0]$. We then tested for pairwise differences between tau loadings. For example, $H = [0, -1, 1, 0, 0, 0, 0]$ tested the differences between lateral temporal tau and medial temporal tau loadings; $H = [0, -1, 0, 1, 0, 0, 0]$ compared posterior cortical tau and medial temporal tau loadings; $H = [0, 0, -1, 1, 0, 0, 0]$ compared posterior cortical tau and lateral temporal tau loadings.

S8. Comparison with canonical correlation analysis (CCA)

To apply CCA to the 149 ADNI-GO/2 participants, for each voxel of the log-transformed GM density images, z transformation (using the mean of CN participants and variance of all participants) was applied. The same procedure was applied to the cognitive scores. To avoid overfitting the CCA, we followed Smith and colleagues' (Smith *et al.*, 2015) procedure of applying PCA to reduce the dimensionality of both atrophy and cognitive scores, retaining 95% of the variance. Varying the number of PCA components did not significantly affect the results. Statistical significance of the CCA modes was established using permutation testing (Smith *et al.*, 2015). In addition, we also performed 10-fold cross-validation, just like in MMLDA.

References

Barnard K, Duygulu P, Forsyth D, de Freitas N, Blei DM, Jordan MI. Matching Words and Pictures. J. Mach. Learn. Res. 2003; 3: 1107–35.

Blei DM, Jordan MI. Modeling annotated data. Proc. 26th Annu. Int. ACM SIGIR Conf. Res. Dev. Inf. Retr. 2003: 127–34.

Blei DM, Ng AY, Jordan MI. Latent Dirichlet Allocation. J. Mach. Learn. Res. 2003; 3: 993–1022.

Putthividhya D, Attias HT, Nagarajan SS, Lee T-W. Probabilistic Graphical Model for Auto-Annotation, Content-Based Retrieval, and Classification of TV Clips Containing Audio, Video, and Text. In ICASSP 2007; 2: 789–92.

Smith SM, Nichols TE, Vidaurre D, Winkler AM, J Behrens TE, Glasser MF, et al. A positive-negative mode of population covariation links brain connectivity, demographics and behavior. Nat. Neurosci. 2015; 18: 1–7.

Zhang X, Mormino EC, Sun N, Sperling RA, Sabuncu MR, Yeo BTT. Bayesian model reveals latent atrophy factors with dissociable cognitive trajectories in Alzheimer's disease. Proc. Natl. Acad. Sci. 2016; 113: E6535-44.

Supplementary Figures and Tables

Behavioral Description	Fieldname	Behavioral Description	Fieldname
ADAS		**MMSE**	
ADAS: Word Recall	Q1SCORE	MMSE: Orientation to time	MMDATE
ADAS: Commands	Q2SCORE		MMYEAR
ADAS: Construction	Q3SCORE		MMMONTH
ADAS: Delayed Word Recall	Q4SCORE		MMDAY
ADAS: Naming	Q5SCORE		MMSEASON
ADAS: Ideational Praxis	Q6SCORE	MMSE: Orientation to place	MMHOSPIT
ADAS: Orientation	Q7SCORE		MMFLOOR
ADAS: Word Recognition	Q8SCORE		MMCITY
ADAS: Recall Instructions	Q9SCORE		MMAREA
ADAS: Spoken Language	Q10SCORE		MMSTATE
ADAS: Word Finding	Q11SCORE	MMSE: Immediate Recall	MMBALL
ADAS: Comprehension	Q12SCORE		MMFLAG
ADAS: Number Cancellation	Q13SCORE		MMTREE
NEUROBAT		MMSE: Attention	MMD
BNT: without cue and semantic cue	BNTTOTAL		MML
CFT: animal	CATANIMSC		MMO
Clock Drawing Test	CLOCKSCOR		MMW
Clock Copying Test	COPYSCOR	MMSE: Delayed Recall	MMBALLDL
LM: Immediate Recall	LIMMTOTAL		MMFLAGDL
LM: Delayed Recall	LDELTOTAL		MMTREEDL
RAVLT: Trial 6	AVTOT6	MMSE: Language	MMWATCH
RAVLT: 30 min delay	AVDEL30MIN		MMPENCIL
RAVLT: recognition	AVDELTOT	MMSE: Repetition	MMREPEAT
TMT: Part A	TRAASCOR	MMSE: Complex commands	MMHAND
TMT: Part B	TRABSCOR		MMFOLD
			MMONFLR
			MMREAD
			MMWRITE
			MMDRAW

Table S1. Chosen cognitive scores from ADAS, MMSE and NEUROBAT spreadsheets. The left column lists the behavioral description and the right column corresponds to the fieldname of the behavior in the spreadsheet. For the MMSE, the blocks show how we merge the item scores into subscores. For example, "MMSE: Orientation to time" was obtained by summing "MMDATE", "MMYEAR", "MMMONTH", "MMDAY" and "MMSEASON". ADAS = Alzheimer's Disease Assessment Scale; MMSE = Mini-Mental State Examination; NEUROBAT = Neuropsychological Battery; BNT = Boston Naming Test; CFT = Category Fluency Test; LM = Logical Memory; RAVLT = Rey Auditory Verbal Learning Test; TMT = Trail Making Test.

Behavioral Description	Behavioral Abbreviation
ADAS: Word Recall	ADAS: Word Recall
ADAS: Commands	ADAS: Commands
ADAS: Construction	ADAS: Construction
ADAS: Delayed Word Recall	ADAS: Del. Recall
ADAS: Naming + MMSE: Language	ADAS: Nam. + MMSE: Lan.
ADAS: Ideational Praxis	ADAS: Ide. Praxis
ADAS: Orientation	ADAS: Orientation
ADAS: Word Recognition	ADAS: Word Recog.
ADAS: Spoken Language + Comprehension	ADAS: Spo. Com.
ADAS: Word Finding	ADAS: Word Find.
ADAS: Number Cancellation	ADAS: Num. Can.
MMSE: Orientation to time	MMSE: Ori. Time
MMSE: Orientation to place	MMSE: Ori. Place
MMSE: Immediate Recall + Delayed Recall	MMSE: Imm. Del.
MMSE: Attention	MMSE: Attention
MMSE: Repetition	MMSE: Repetition
MMSE: Complex Commands	MMSE: Com. Com.
BNT: without cue and semantic cue	BNT
CFT: animal	CFT: Animal
Clock Drawing Test	Clock Drawing Test
Clock Copying Test	Clock Copying Test
LM: Immediate Recall	LM: Imm. Recall
LM: Delayed Recall	LM: Del. Recall
RAVLT: Trail 6	RAVLT: Trail 6
RAVLT: 30 min delay	RAVLT: 30 min delay
RAVLT: Recognition	RAVLT: Recognition
TMT: Part A	TMT: Part A
TMT: Part B	TMT: Part B

Table S2. 28 cognitive scores after combination and corresponding abbreviations. The left column lists combined scores from Table S1 and right column corresponds to abbreviations of these scores shown in Fig. 4.

	MTL-Memory	Lateral Temporal-Language	Posterior Cortical-Executive	Overall p*
Baseline age (years)	75.4 (7.7)	74.9 (7.6)	73.4 (8.8)	**2e-4**
Education (years)	15.9 (2.7)	15.8 (2.7)	15.6 (2.6)	0.44
Sex (0 for male)	0.4 (0.5)	0.4 (0.5)	0.4 (0.5)	0.91
Amyloid (SUVR)‡	1.4 (0.2)	1.4 (0.2)	1.4 (0.2)	0.16
APOE ε2§	0.06 (0.3)	0.07 (0.3)	0.06 (0.3)	0.97
APOE ε4§	0.87 (0.7)	0.86 (0.7)	0.89 (0.7)	0.82

Table S3. Characteristics of 149 AD dementia patients by factor. Data are weighted averages (weighted standard deviation) with weights corresponding to factor probabilities. Highlighted p values (blue) are characteristics significantly different across factors. *Computed by linear hypothesis test on GLM or likelihood ratio test on logistic regression for sex (see Methods). ‡Only available for 146 patients. §The original counts were 0, 1 or 2. Only available for 145 patients.

ADNI-1 CN (N = 222)		ADNI-GO/2 CN (N = 292)	
Cognitive Scores	Standard Deviation	Cognitive Scores	Standard Deviation
ADAS: Recall Instructions	0	MMSE: Immediate Recall	0
MMSE: Language	0	MMSE: Language	0.058521
ADAS: Comprehension	0.067116	ADAS: Spoken Language	0.14211
MMSE: Immediate Recall	0.094701	ADAS: Recall Instructions	0.15415
ADAS: Spoken Language	0.13332	ADAS: Comprehension	0.21664
ADAS: Ideational Praxis	0.2175	ADAS: Naming	0.24707
ADAS: Naming	0.25157	ADAS: Ideational Praxis	0.27887
MMSE: Repetition	0.26652	MMSE: Orientation to time	0.31291
ADAS: Word Finding	0.30836	MMSE: Repetition	0.31291
MMSE: Orientation to time	0.30836	ADAS: Word Finding	0.35018
ADAS: Orientation	0.33905	ADAS: Orientation	0.35193
MMSE: Orientation to place	0.34607	Clock Copying Test	0.36978
ADAS: Commands	0.36457	MMSE: Orientation to place	0.3802
MMSE: Complex commands	0.38148	MMSE: Complex commands	0.38225
Clock Copying Test	0.42899	ADAS: Commands	0.38745
ADAS: Construction	0.50089	ADAS: Construction	0.52695
MMSE: Attention	0.51436	MMSE: Delayed Recall	0.52812
MMSE: Delayed Recall	0.55441	Clock Drawing Test	0.601
Clock Drawing Test	0.66811	MMSE: Attention	0.64344
ADAS: Number Cancellation	0.68254	ADAS: Number Cancellation	0.64931
ADAS: Word Recall	1.1443	ADAS: Word Recall	1.3187
ADAS: Delayed Word Recall	1.7116	ADAS: Word Recognition	1.7797
ADAS: Word Recognition	2.2864	ADAS: Delayed Word Recall	1.8982
BNT: without cue and semantic cue	2.3108	BNT: without cue and semantic cue	2.0933
RAVLT: recognition	2.526	RAVLT: recognition	2.6186
RAVLT: Trial 6	3.4189	LM: Immediate Recall	3.0363
LM: Immediate Recall	3.4926	LM: Delayed Recall	3.0429
LM: Delayed Recall	3.5898	RAVLT: Trial 6	3.4539
RAVLT: 30 min delay	3.6879	RAVLT: 30 min delay	4.0232
CFT: animal	5.5864	CFT: animal	5.2792
TMT: Part A	13.0895	TMT: Part A	11.283
TMT: Part B	44.611	TMT: Part B	42.7338

Table S4. **Standard deviation of cognitive scores in ADNI-1 and ADNI-GO/2 CN participants.**

ADNI-1 CN (N = 222)		ADNI-GO/2 CN (N = 292)	
Cognitive Scores	Standard Deviation	Cognitive Scores	Standard Deviation
ADAS:Spoken Language+Comprehension	0.14871	ADAS:Naming+MMSE:Language	0.25302
ADAS:Naming+MMSE:Language	0.25157	ADAS:Spoken Language+Comprehension	0.28189
MMSE:Immediate Recall+Delayed Recall	0.5654	MMSE:Immediate Recall+Delayed Recall	0.52812

Table S5. Standard deviation of combined cognitive scores in ADNI-1 and ADNI-GO/2 CN participants.

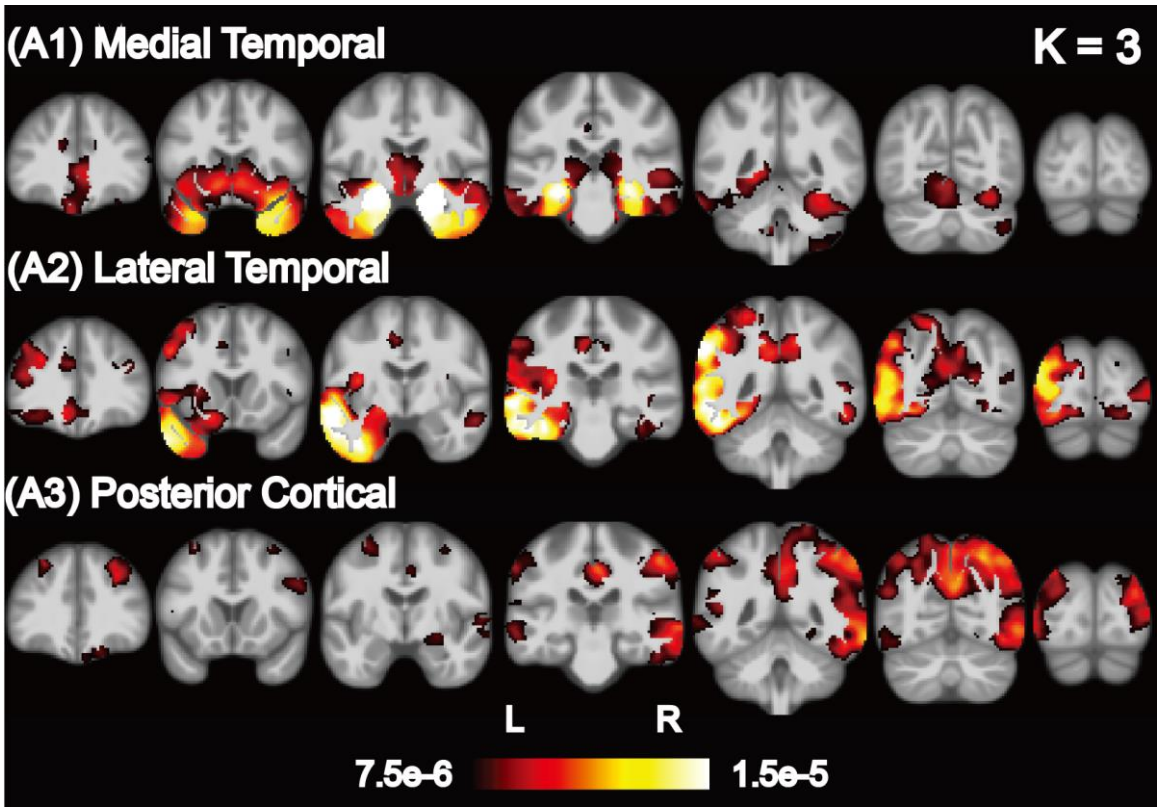

Figure S1. Probabilistic atrophy maps of K = 3 latent factors in ADNI-GO/2. Brighter color indicates higher probability of atrophy at that voxel for a particular factor (i.e., Pr(Voxel | Factor)). Each of the three factors was associated with a distinct pattern of brain atrophy.

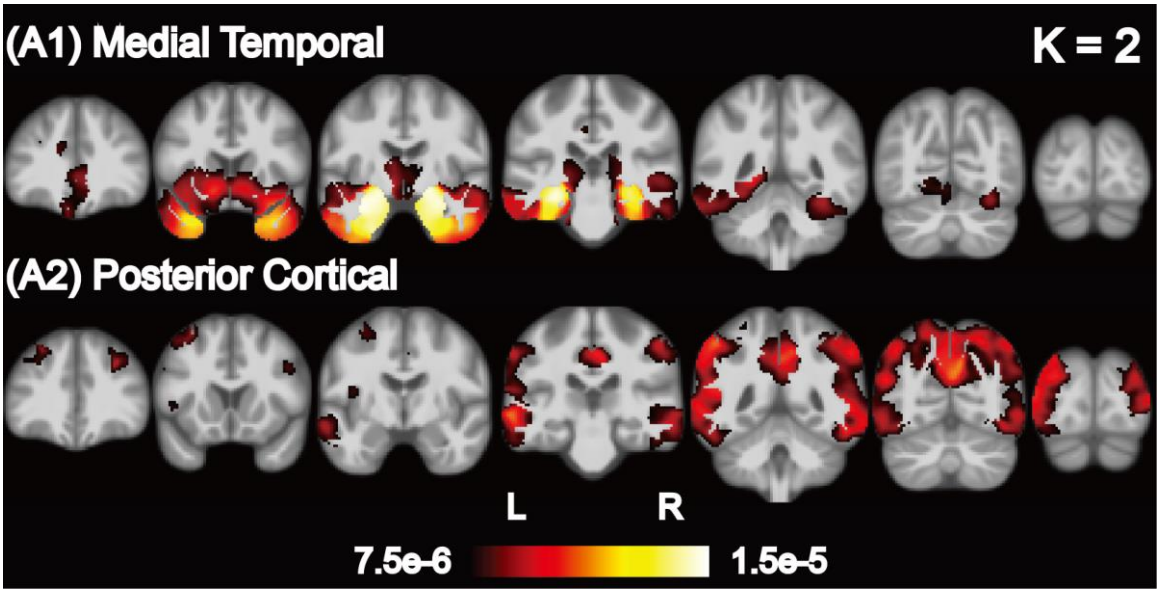

Figure S2. Probabilistic atrophy maps for K = 2 latent factors in ADNI-GO/2. Brighter color indicates higher probability of atrophy at that voxel for a particular factor (i.e., Pr(Voxel | Factor)).

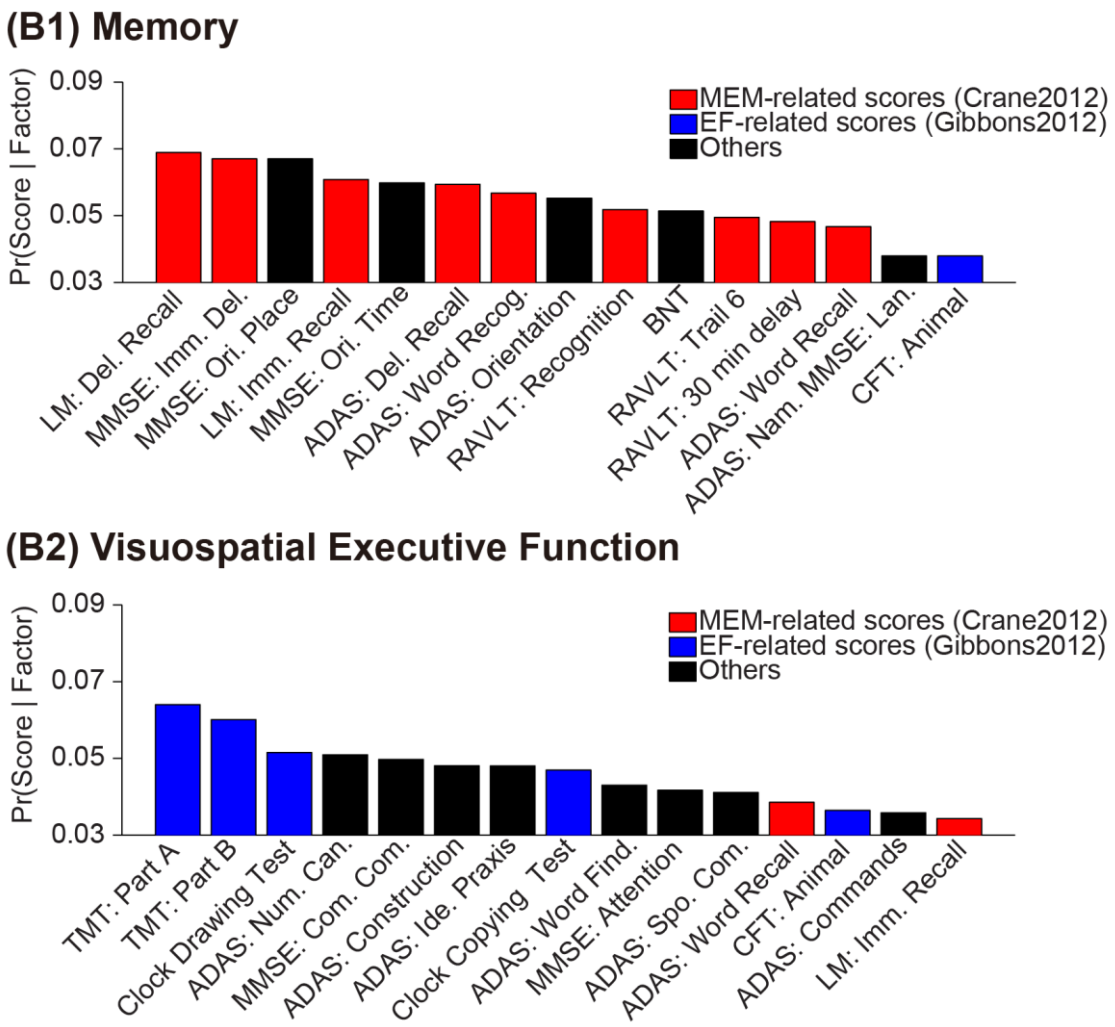

Figure S3. Probabilistic cognitive deficits for K = 2 latent factors in ADNI-GO/2. Red bars indicate scores associated with memory, blue bars indicate scores associated with executive function and black bars indicate scores that do not belong to either memory or executive function. The probabilities of cognitive scores associated with a factor (i.e., Pr(Score | Factor)) are sorted from high to low and only top 15/28 cognitive scores are shown in the plot.

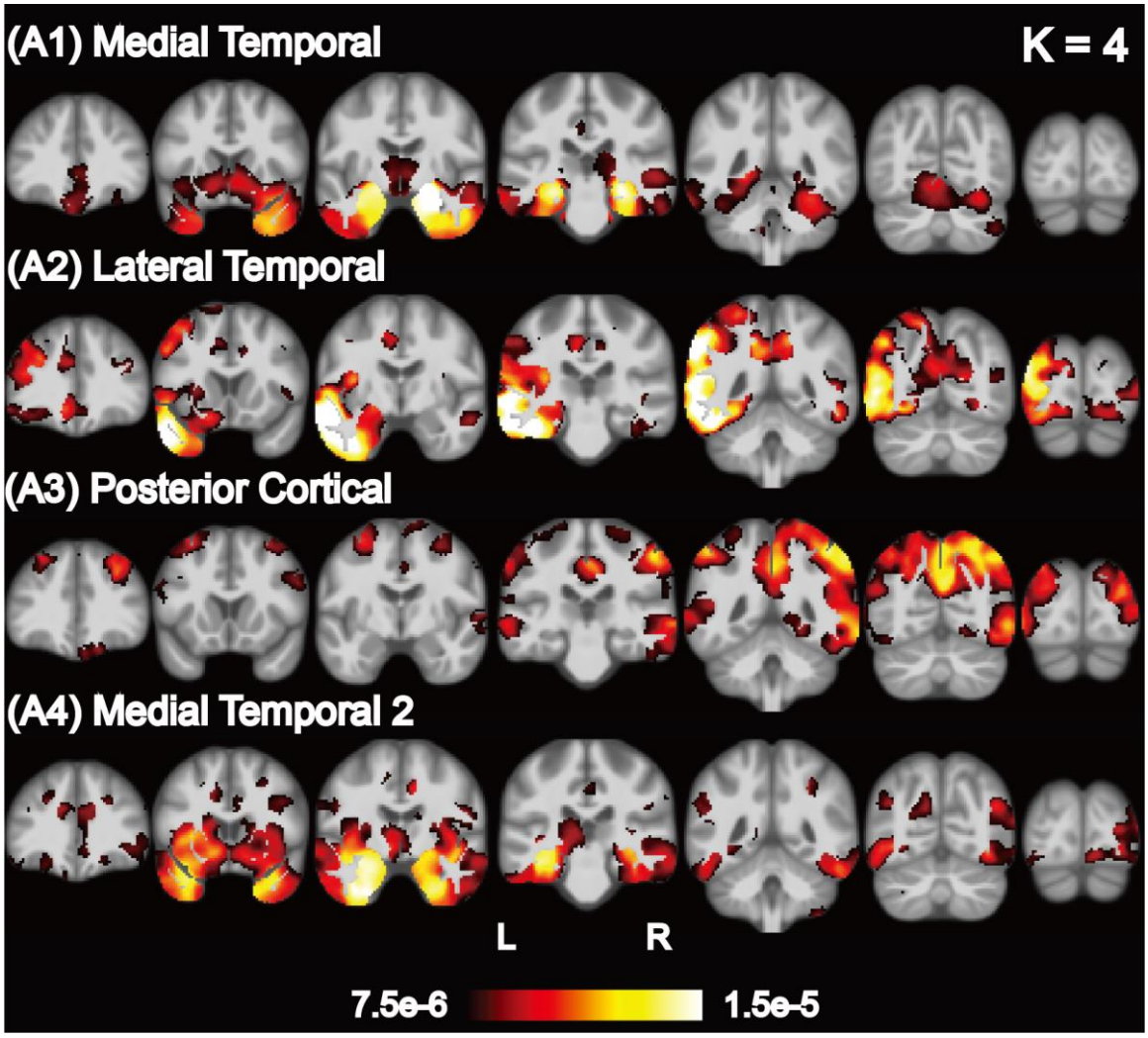

Figure S4. Probabilistic atrophy maps for K = 4 latent factors in ADNI-GO/2. Brighter color indicates higher probability of atrophy at that voxel for a particular factor (i.e., Pr(Voxel | Factor)).

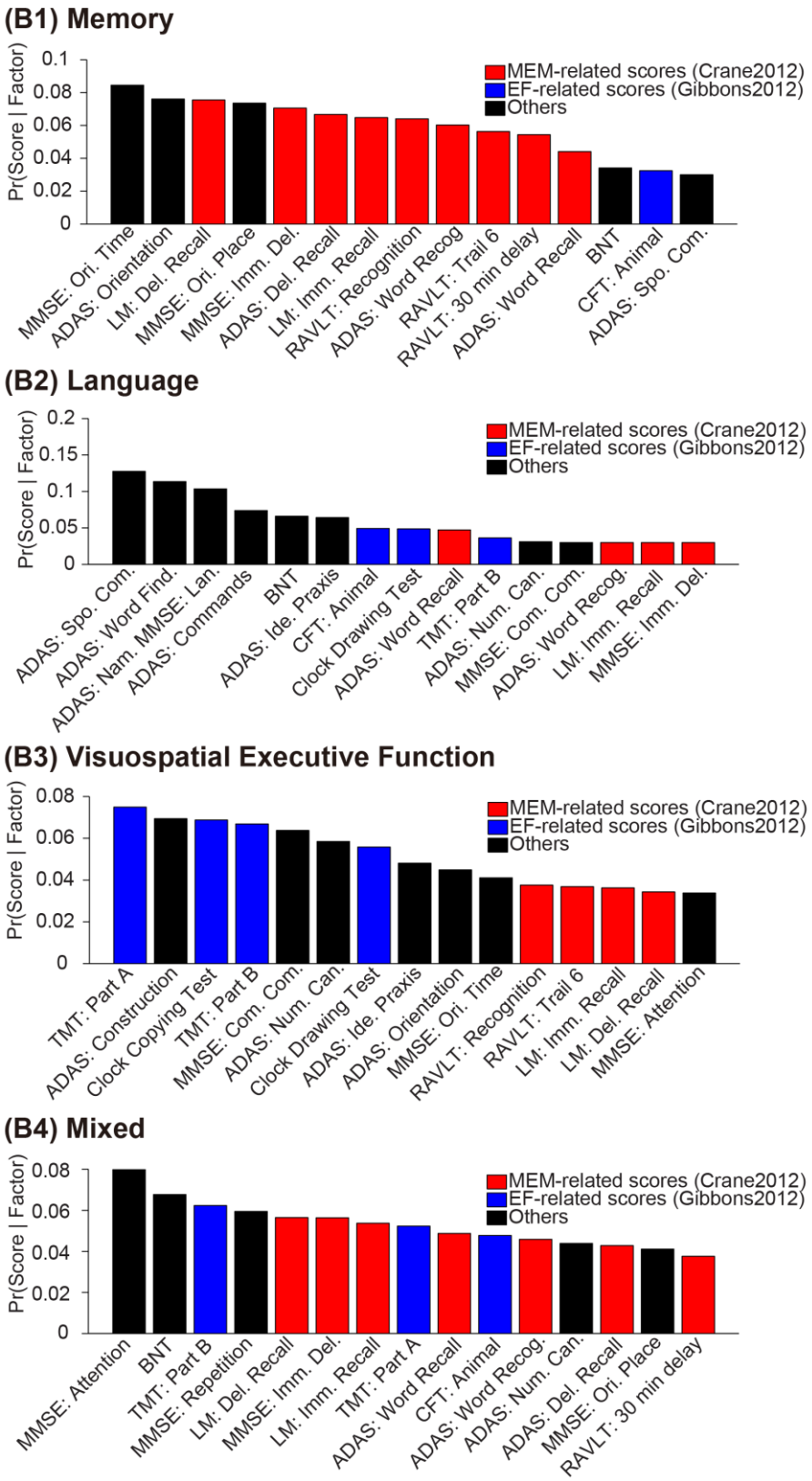

Figure S5. Probabilistic cognitive deficits for K = 4 latent factors in ADNI-GO/2.

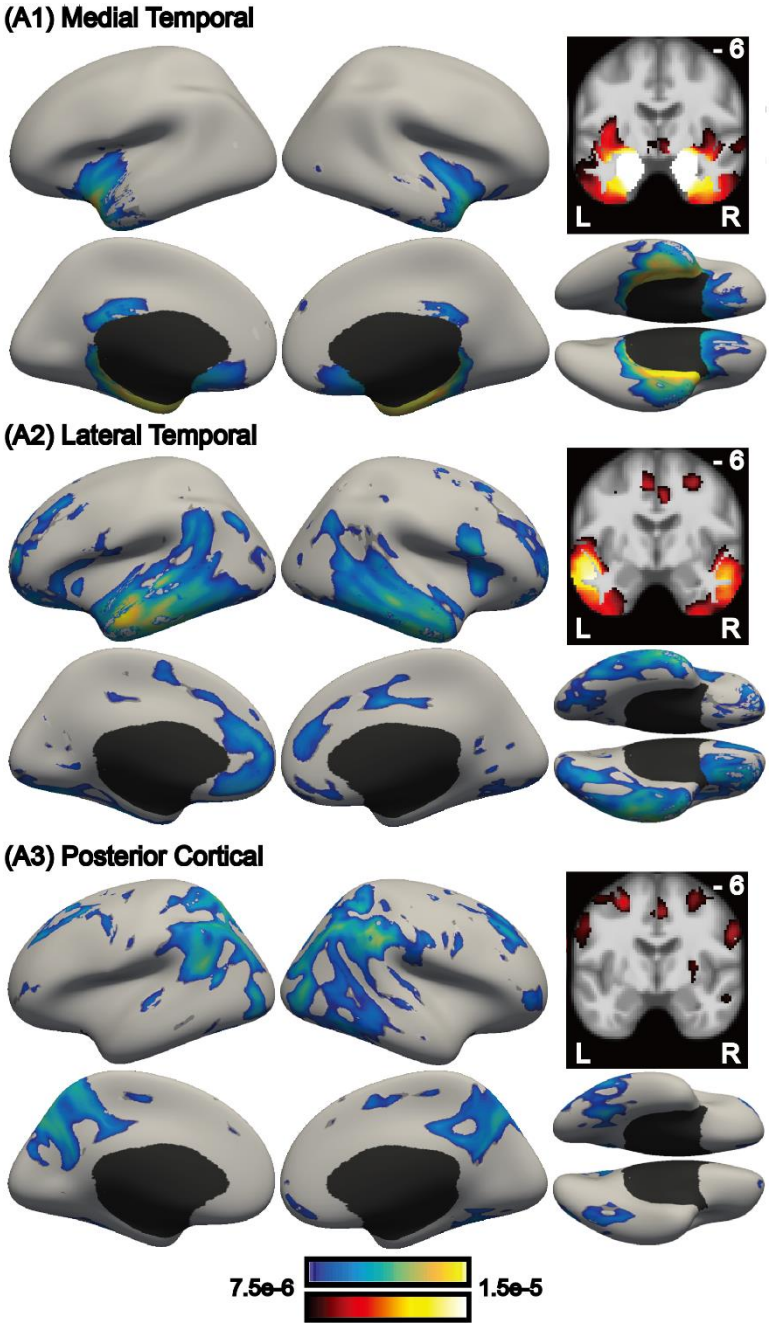

Figure S6. Probabilistic atrophy maps for K = 3 latent factors in ADNI-1. Brighter color indicates higher probability of atrophy at that voxel for a particular factor (i.e., Pr(Voxel | Factor)). Each of the three factors was associated with a distinct pattern of brain atrophy.

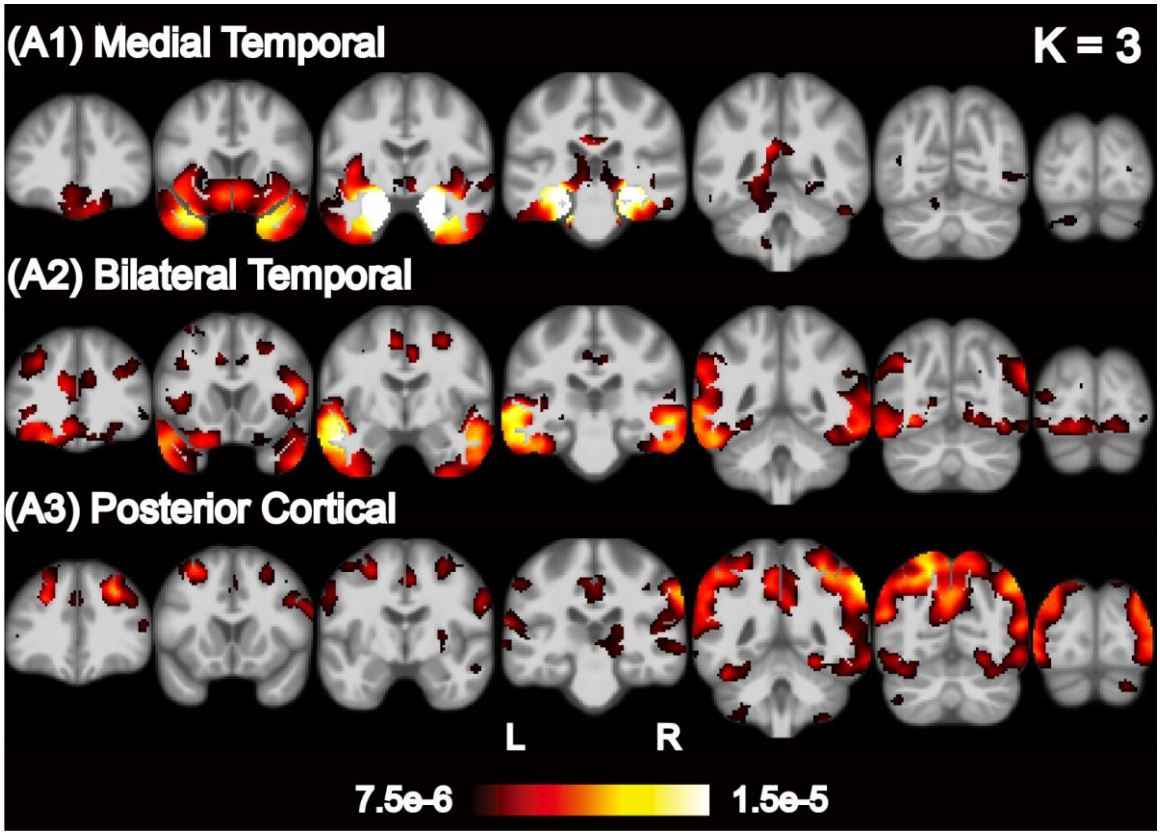

Figure S7. Probabilistic atrophy maps for K=3 latent factors in ADNI-1. Bright color indicates higher probability of atrophy at that spatial location for a particular atrophy factor (i.e., Pr(Voxel | Factor)). Each of the three factors was associated with a distinct pattern of brain atrophy.

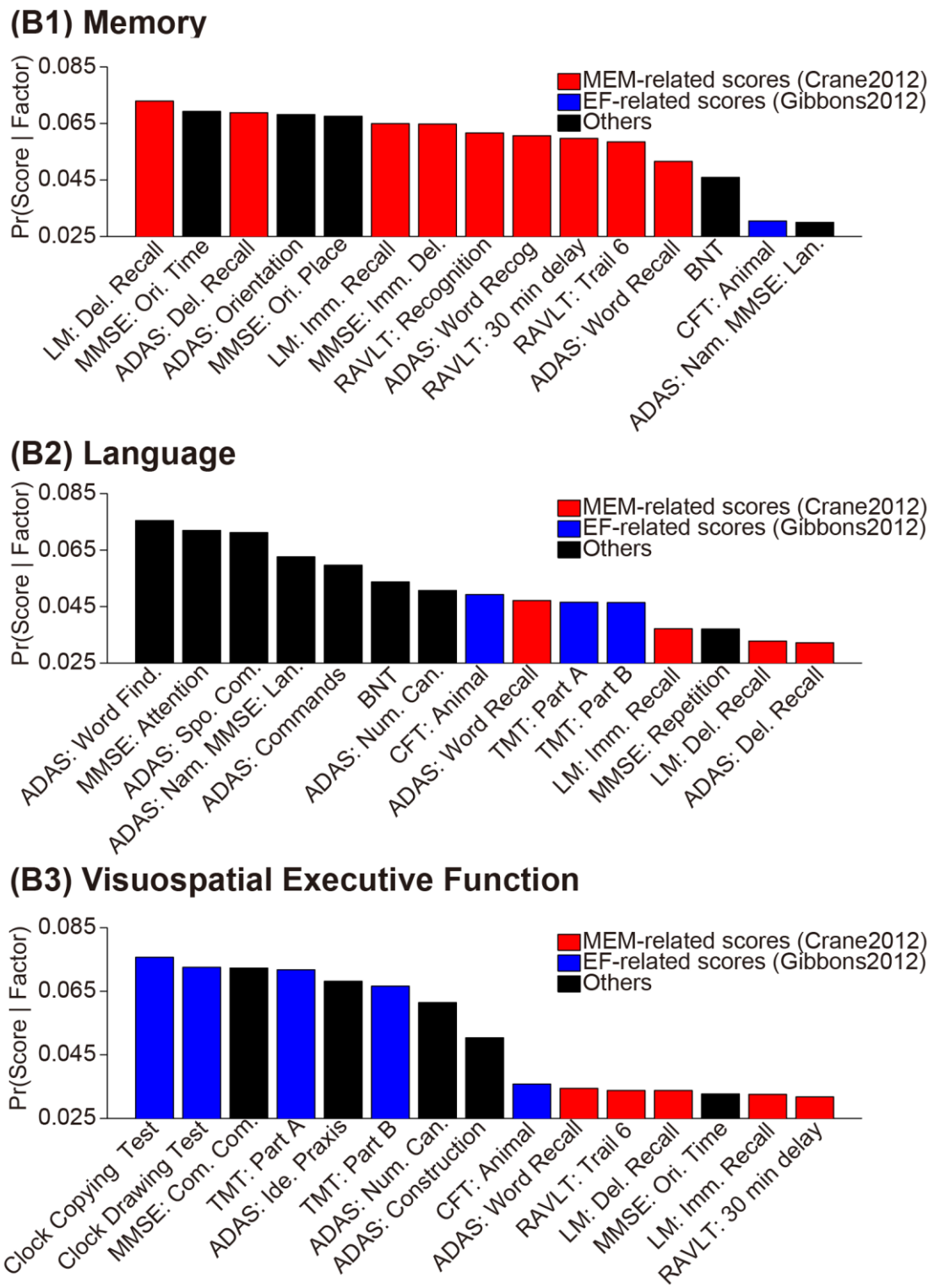

Figure S8. Probabilistic cognitive deficits for K=3 latent factors in ADNI-1. Red bars indicate scores associated with memory, blue bars indicate scores associated with executive function and black bars indicate scores that do not belong to either memory or executive function. The probabilities of cognitive scores associated with a factor (i.e., Pr(Score | Factor)) are sorted from high to low and only top 15/28 cognitive scores are shown in the plot.

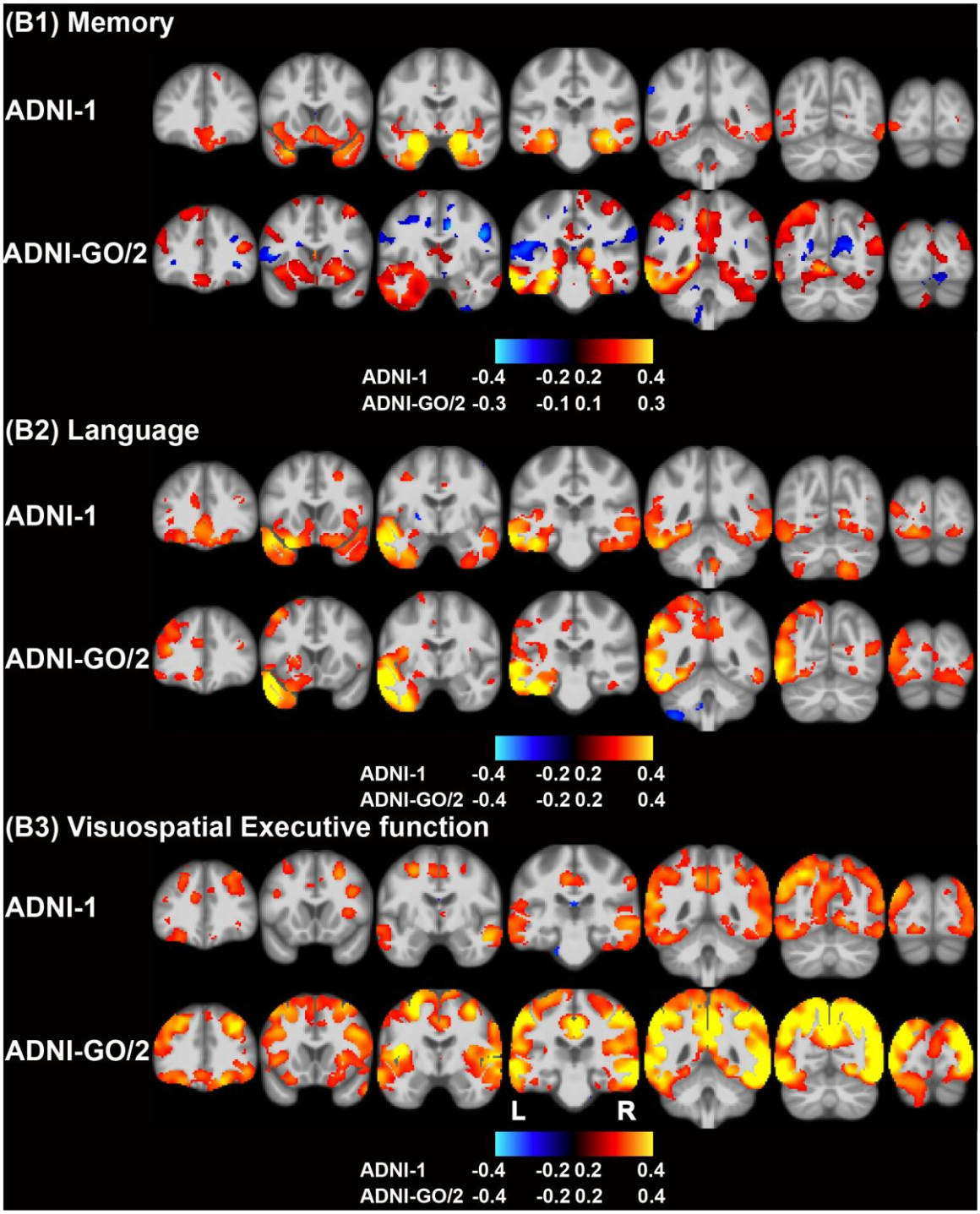

Figure S9. Correlation maps between atrophy z score and cognitive mean z score of top 5 scores (Fig. 3) for each factor in ADNI-1 or ADNI-GO/2 AD. Hot color indicates positive correlation and cold color indicates negative correlation.

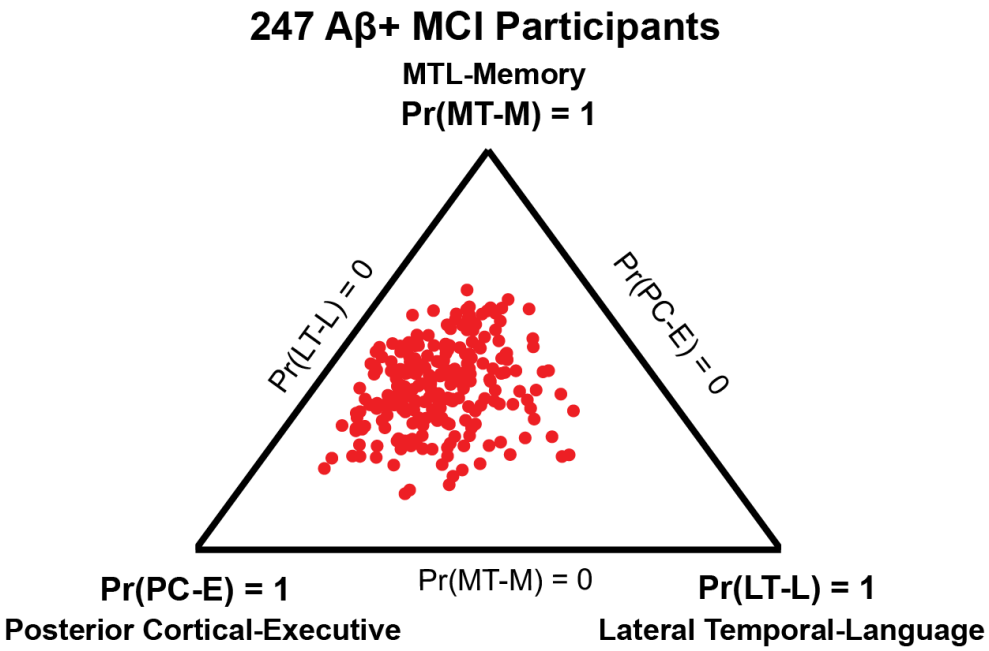

Figure S10. Factor compositions of 247 Aβ+MCI participants for K = 3 factors. Each participant corresponds to a dot, whose location (in barycentric coordinates) represents the factor composition. Corners of the triangle represent "pure factors"; closer distance to the respective corners indicates higher probabilities for the respective factors. Most dots are far from the corners, suggesting that most participants expressed multiple factors. "MT-M" indicates the MTL-Memory factor. "PC-E" indicates the Posterior Cortical-Executive factor. "LT-L" indicates the Lateral Temporal-Language factor.

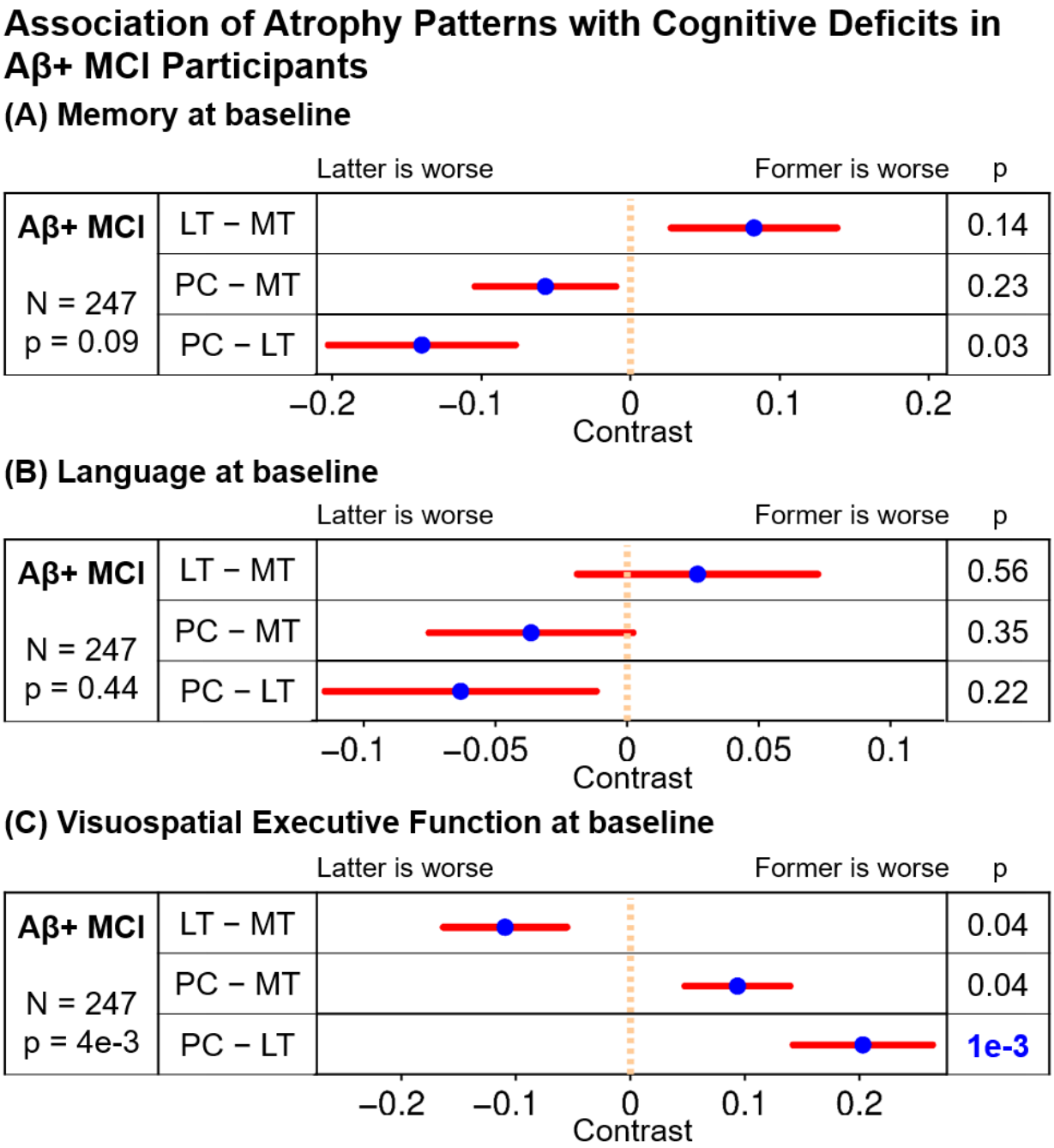

Figure S11. Associations between atrophy loadings and cognitive loadings among Aβ+ MCI participants. All p values remaining significant after FDR (q = 0.05) control are highlighted in blue. MT, LT, and PC indicate medial temporal lobe, lateral temporal, and cortical atrophy patterns, respectively. Blue dots are estimated differences in regression coefficients between atrophy loadings and red bars show standard errors.

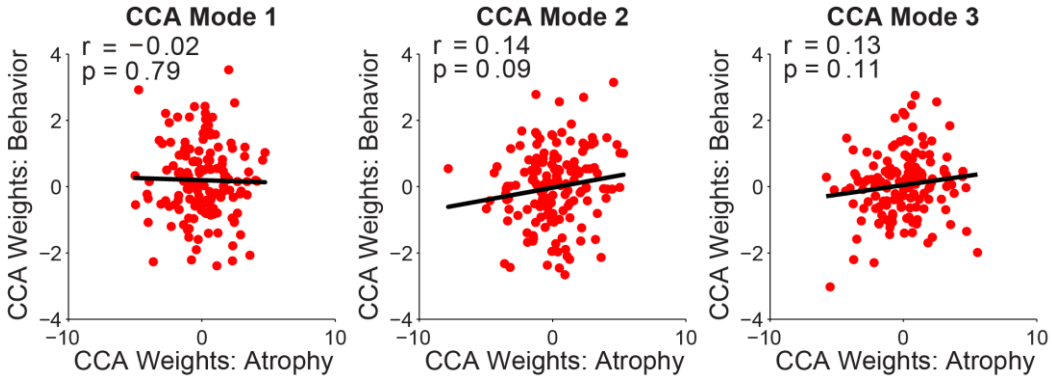

Figure S12. 10-fold cross validation of three CCA modes. We performed 10-fold cross-validation among the AD dementia patients in ADNI-GO/2. Each red dot represents an AD patient.

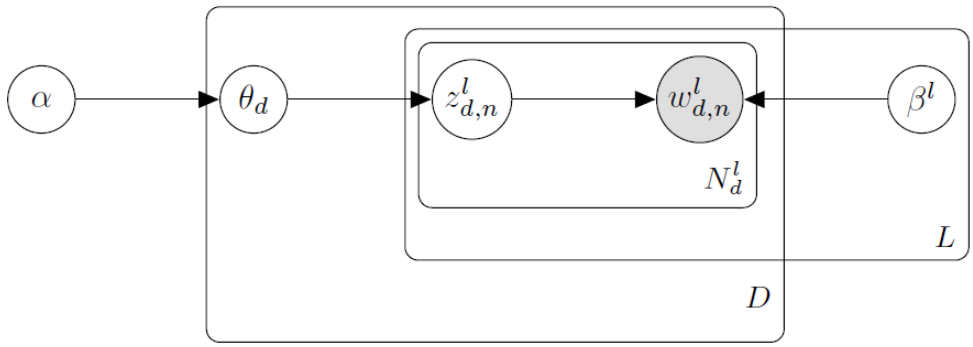

Figure S13. Graphical model representation of the MMLDA model. The circled variables represent random variables, while the shaded node represents observed variables. The edges represent statistical dependencies. The variables at the corner of the rectangles indicate the number of times the variables inside the rectangles were replicated. Therefore θ_d was replicated D times, once for each participant. The variables $z^l_{d,n}$, $w^l_{d,n}$ and β^l were replicated L times, once for each modality. For the d-th participant, the variables $z^l_{d,n}$ and $w^l_{d,n}$ were replicated N^l_d times, once for one unit in an atrophy voxel or a cognitive score. α is hyperparameter parameterizing the Dirichlet prior on θ. θ denote Pr(Factor | Participant), β^1 denote Pr(Voxel | Factor) and β^2 denote Pr(Score | Factor) . Therefore, θ, β^1 and β^2 are matrices, where each row is a categorical distribution summing to one. $z^l_{d,n}$ denote the factor that the voxel or score belongs to and $w^l_{d,n}$ denote the location of observed atrophy voxel or cognitive score.

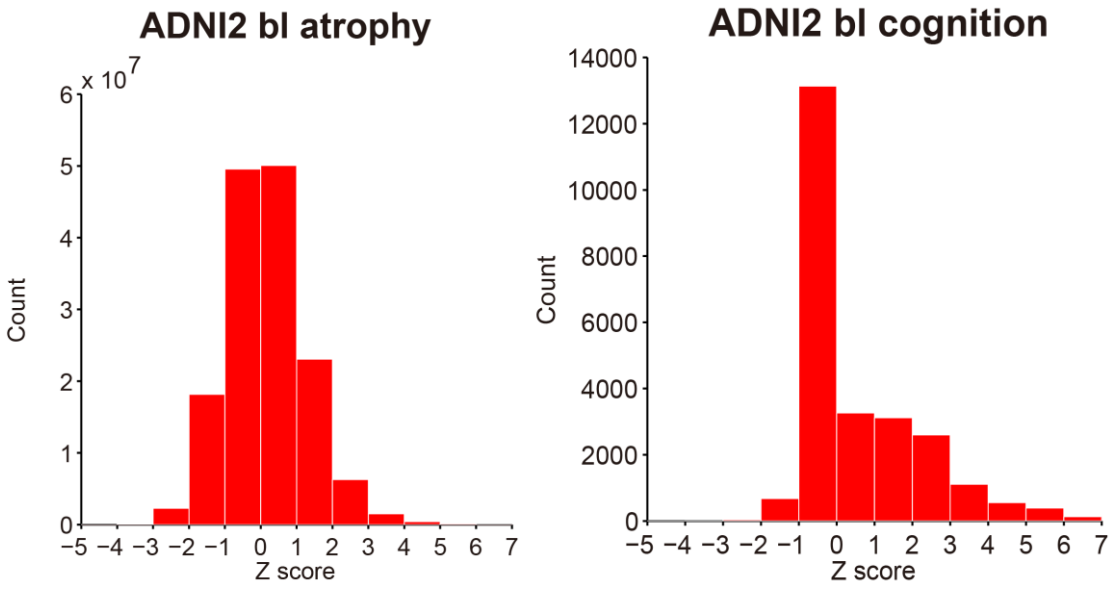

Figure S14. Histogram of atrophy z score and cognitive z score in ADNI-2. The larger z score corresponds to worse performance.