Heterogeneity of neuroanatomical patterns in prodromal Alzheimers: links to cognition, progression, biomarkers

Aoyan Dong, Jon B. Toledo, Nicolas Honnorat, Jimit Doshi, Erdem Varol, Aristeidis Sotiras, David Wolk, John Q. Trojanowski, Christos Davatzikos; for the Alzheimers Disease Neuroimaging Initiative

Methods

Methodological details of the semi-supervised clustering approach

This clustering method treats subjects as points in a high-dimensional feature space, where both the normal control group and the patients are viewed as point distributions. As shown in **Supplementary Fig. 1**, the point distribution has two sources of variations: the population variation and covariate effects. The covariates could be factors like age and gender, and in this study, we added the study cohort ID as one covariate. The standard clustering methods group the patients based on the distances between individuals, which results in clustering the largest factor of data variability instead of the disease effect. In order to address this challenge, the probabilistic approach proposed in (Dong *et al.*, 2016) considers the disease effect as a transformation from the normal control distribution to the patient distribution. The data variability introduced by covariates other than the disease effects were explicitly controlled in the distribution mapping.

Let us assume that there are M normal control subjects $\mathbf{X} = \{x_1, , x_M\}$ and N patient subjects $\mathbf{Y} = \{y_1, , y_N\}$. They are described by two sets of features: a set of D_1-dimensional imaging features, $x_m^v, y_n^v \in \mathcal{R}^{D_1}$ and a set of D_2-dimensional covariate features, $x_m^c, y_n^c \in \mathcal{R}^{D_2}$. For simplicity, subjects are denoted in compact vector forms: $x_m = (x_m^v, x_m^c)$ and $y_n = (y_n^v, y_n^c)$. The clustering approach minimizes the following energy function:

$$\mathcal{E}(\mathbf{X}, \mathbf{Y}, \mathbf{\Theta}) = -\mathcal{L}(\mathbf{X}, \mathbf{Y}, \Theta) + \mathcal{R}(\mathbf{\Theta})$$

where $\mathbf{\Theta}$ denotes the parameters of the model, such as transformations that are applied to $\mathbf{X}$ in order to generate $\mathbf{Y}$; $\mathcal{L}$ is the log-likelihood of the distributions $\mathbf{X}$ and $\mathbf{Y}$ given the parameters; and $\mathcal{R}$ is a regularization term aiming to improve the stability of the clustering results.

We denote the distribution transformation as $\mathbf{T}$, and we assume that $\mathbf{T}$ can be decomposed into a convex combination of K transformations, each one corresponding to a different disease effect. $\mathbf{T}$ maps the imaging feature of a normal control sample to the patient distribution, while keeping its covariate feature unchanged: $\mathbf{T}(x_m) = (\sum_{k=1}^{K} T_k(x_m^v), x_m^c)$. Each transformed normal control point is then considered as a centroid of a spherical Gaussian cluster, and patient points are treated as independent data generated by a Gaussian mixture model with equal weights for each cluster. The data likelihood of this mixture model is optimized during the distribution matching, where covariate features are embedded in the distance between points with a multi-kernel setting. These model assumptions lead to the log-likelihood term $\mathcal{L}$ being:

$$\mathcal{L}(\mathbf{X}, \mathbf{Y}, \Theta) = \sum_{n=1}^{N} \log \sum_{m=1}^{M} \frac{1}{M} \frac{r^{D_2/2}}{(\sqrt{2\pi}\sigma)^{D_1+D_2}} \exp\left\{ \frac{\|y_n^v - \sum_{k=1}^{K} T_k(x_m^v)\|^2 + r\|y_n^c - x_m^c\|^2}{-2\sigma^2} \right\}.$$

We chose affine transformation in this study, i.e. $T_k(x_m^v) = A_k x_m^v + b_k$. Posing Gaussian priors for the parameters leads to the Frobenius norm of $A_k - I$ and the ℓ_2 norm of b_k to be regularized, where I is the identity matrix.

$$\mathcal{R}(\mathbf{\Theta}) = \lambda_1 \sum_k \|b_k\|_2^2 + \lambda_2 \sum_k \|A_k - I\|_F^2$$

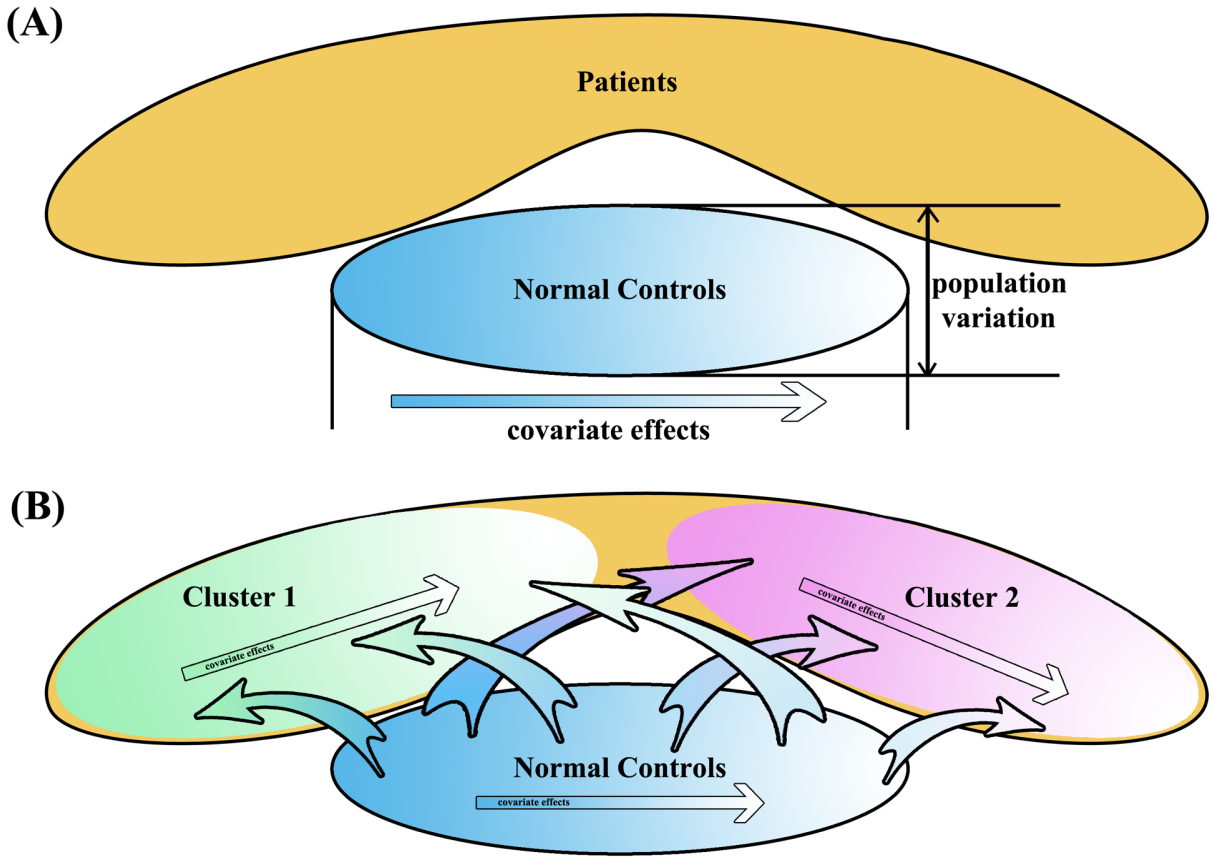
The energy function $\mathcal{E}(\mathbf{X}, \mathbf{Y}, \mathbf{\Theta})$ is optimized with an expectation-maximization approach. The heterogeneous clusters of patients are further generated by the estimated transformations.

The clustering membership used in the manuscript was obtained by running the method multiple times and selecting the most reproducible clustering. In order to evaluate the reproducibility of the memberships, we performed a 10-fold cross validation. The dataset was randomly partitioned into 10 folds. In each fold, transformation parameters were found using subjects in the other nine folds, and subjects in this fold were assigned to clusters according to the transformations. We concatenated training and validation labels as one clustering assignment, and computed the overlap between this assignment and our reported clustering assignment. The average overlap ratio for the 10 folds is 84.1%.

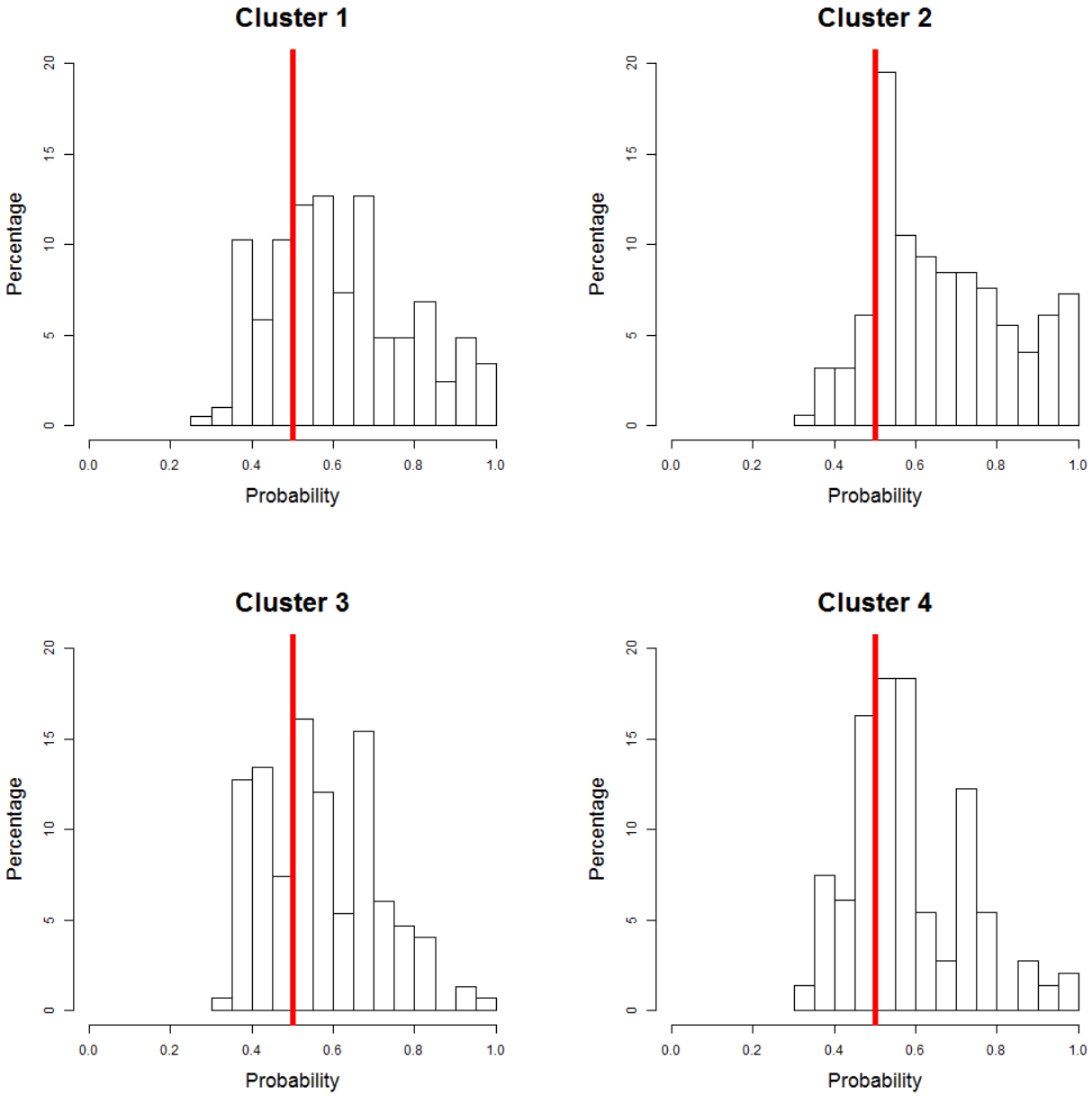
Reference

Dong A, Honnorat N, Gaonkar B, Davatzikos C. CHIMERA: Clustering of Heterogeneous Disease Effects via Distribution Matching of Imaging Patterns. Med. Imaging, IEEE Trans. 2016; 35: 612-621.

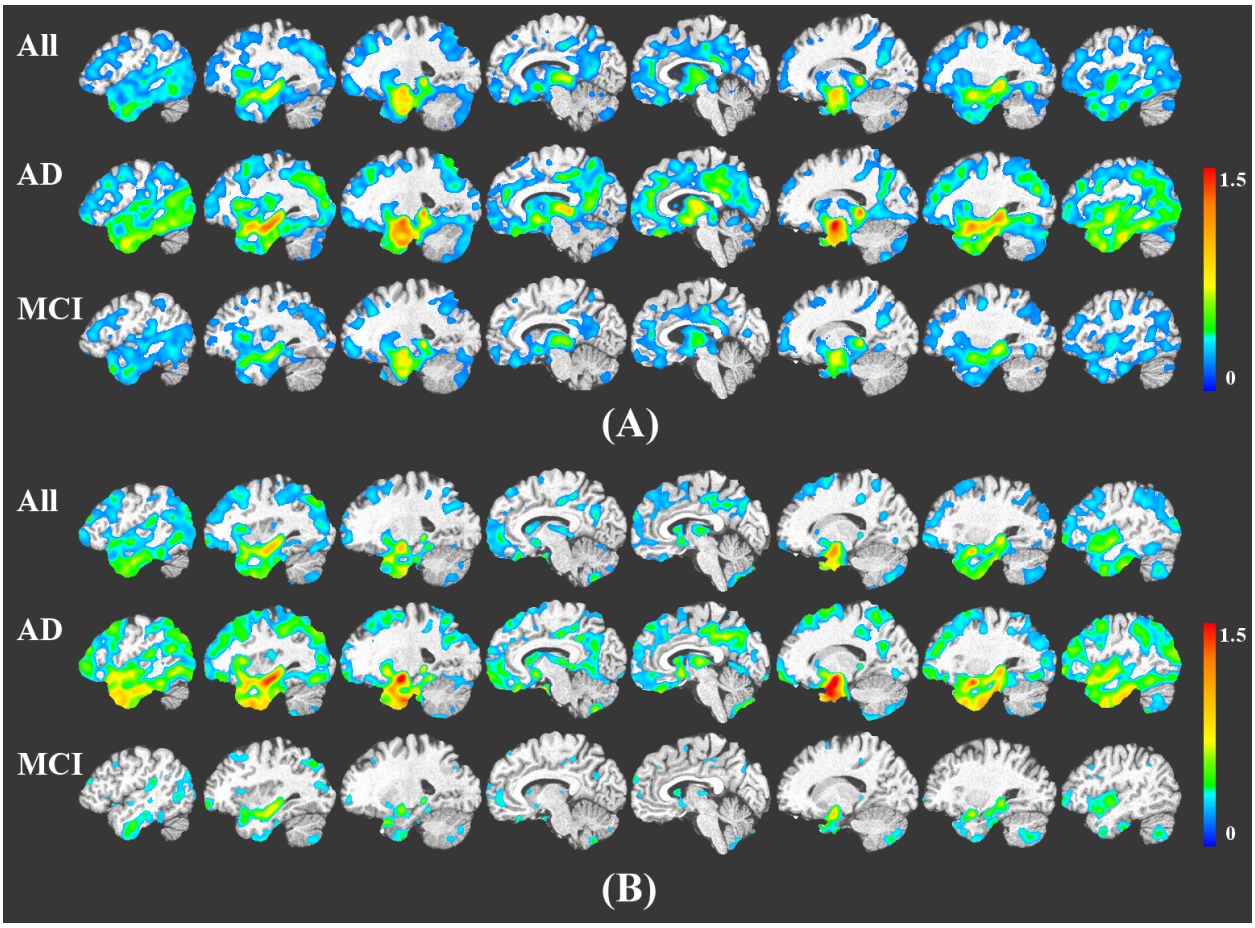
Supplementary Figure 1. Schema of the semi-supervised clustering approach used. The problem assumption illustrated in (A) and the method concept illustrated in (B).



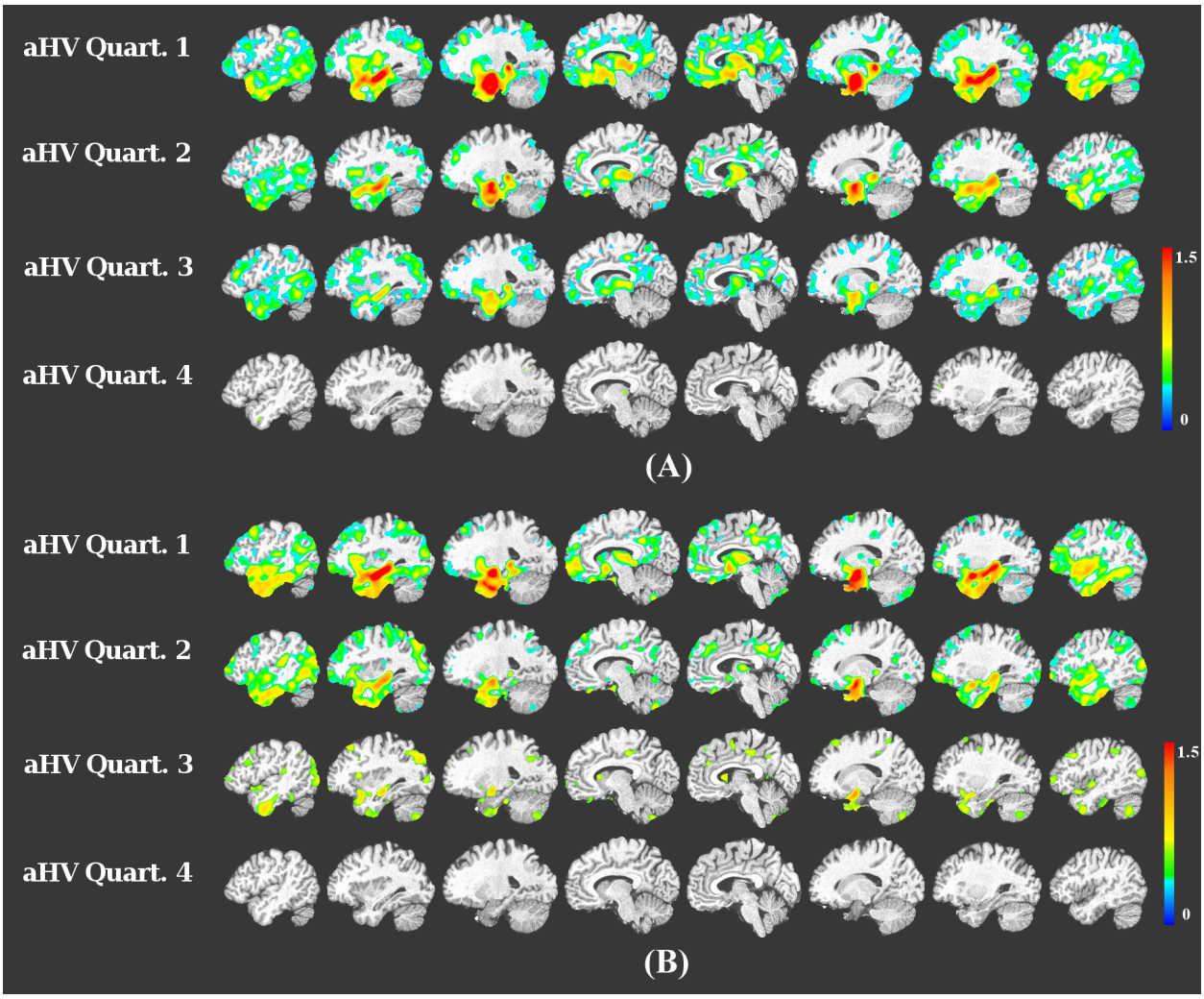
Supplementary Figure 2. Cluster probabilities for subjects classified within each cluster. Vertical red line indicates a 0.5 probability of belonging to the cluster.



Supplementary Figure 3. VBM between clinical groups (MCI and AD) and CN reference group in the ADNI-1 (A) and ADNI-GO/2 cohorts (B). Color scale represents the effect size of gray matter RAVENS maps of each comparison between a cluster and CN individuals. Red indicates greater atrophy (lower volume). Effect size maps are thresholded at false discovery rate (FDR) adjusted p-value of 0.05.



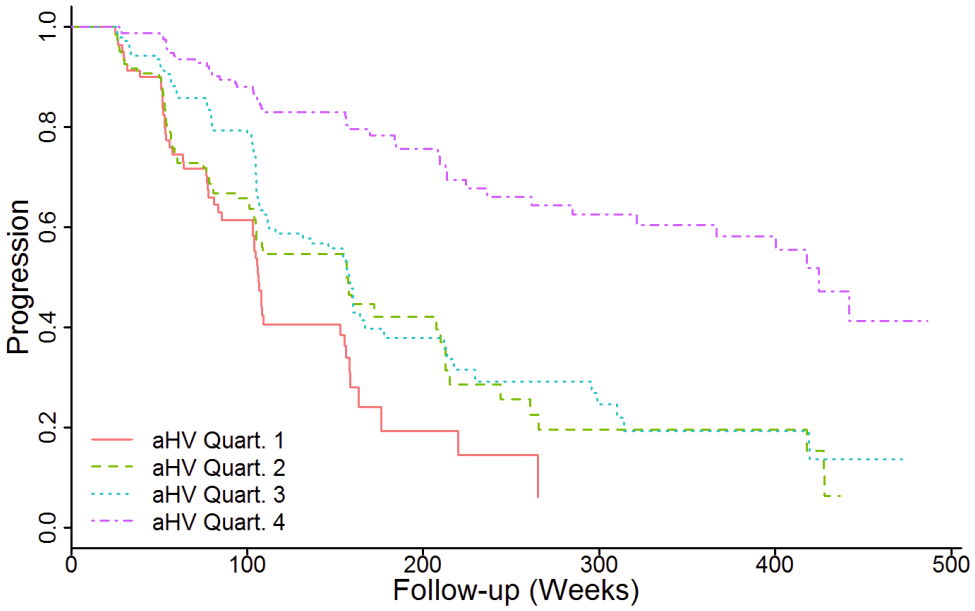
Supplementary Figure 4. VBM between patient clusters stratified by adjusted hippocampal volumes and CN reference group in the ADNI-1 (A) and ADNI-GO/2 cohorts (B). Color scale represents the effect size of gray matter RAVENS maps of each comparison between a cluster and CN individuals. Red indicates greater atrophy (lower volume). Effect size maps are thresholded at false discovery rate (FDR) adjusted p-value of 0.05. Quartile 1 represents lowest volume, whereas Quartile 4 is the highest.



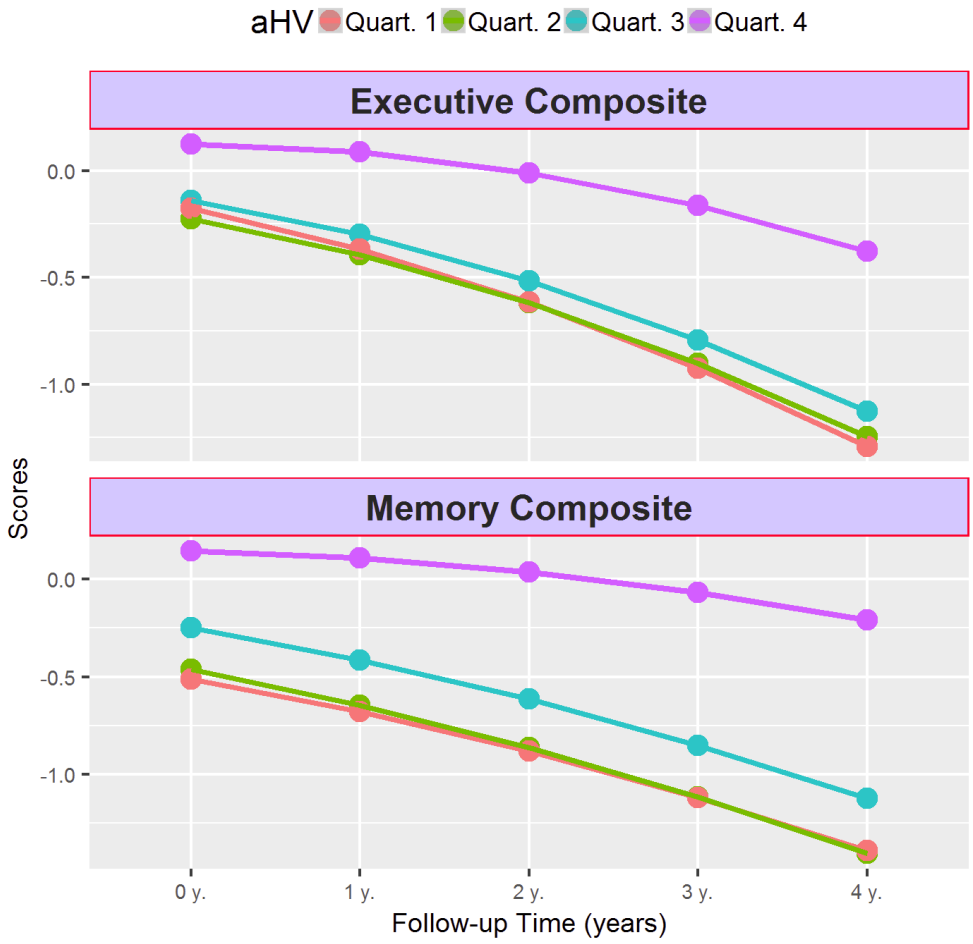
Supplementary Table 1. Regression coefficients and p-values of studying longitudinal associations of cognitive measures with aHV quartiles (Quartile 1 represents lowest volume, whereas Quartile 4 is the highest).

Quartile 1 as reference				
	Memory Composite		Executive Composite	
	Coefficient	p-value	Coefficient	p-value
Quartile 1	Ref.	Ref.	Ref.	Ref.
Quartile 2	-0.130	<0.0001	-0.122	<0.0001
Quartile 3	-0.147	<0.0001	-0.130	<0.0001
Quartile 4	-0.131	<0.0001	-0.154	<0.0001
Quartile 4 as reference				
	Memory Composite		Executive Composite	
	Coefficient	p-value	Coefficient	p-value
Quartile 1	-0.016	<0.0001	0.154	<0.0001
Quartile 2	0.001	0.97	0.032	0.36
Quartile 3	0.131	0.60	0.024	0.51
Quartile 4	Ref.	Ref.	Ref.	Ref.

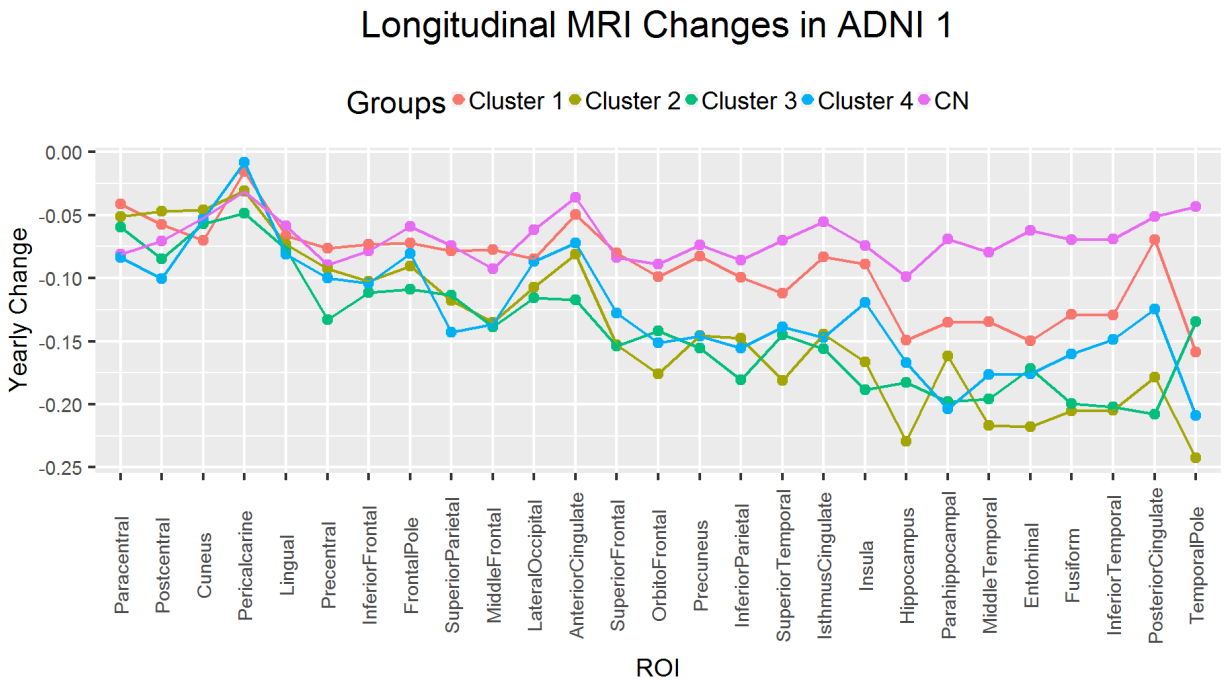
Supplementary Figure 5. Progression from MCI to AD based on aHV quartiles. Quartile 1 represents lowest volume, whereas Quartile 4 is the highest.



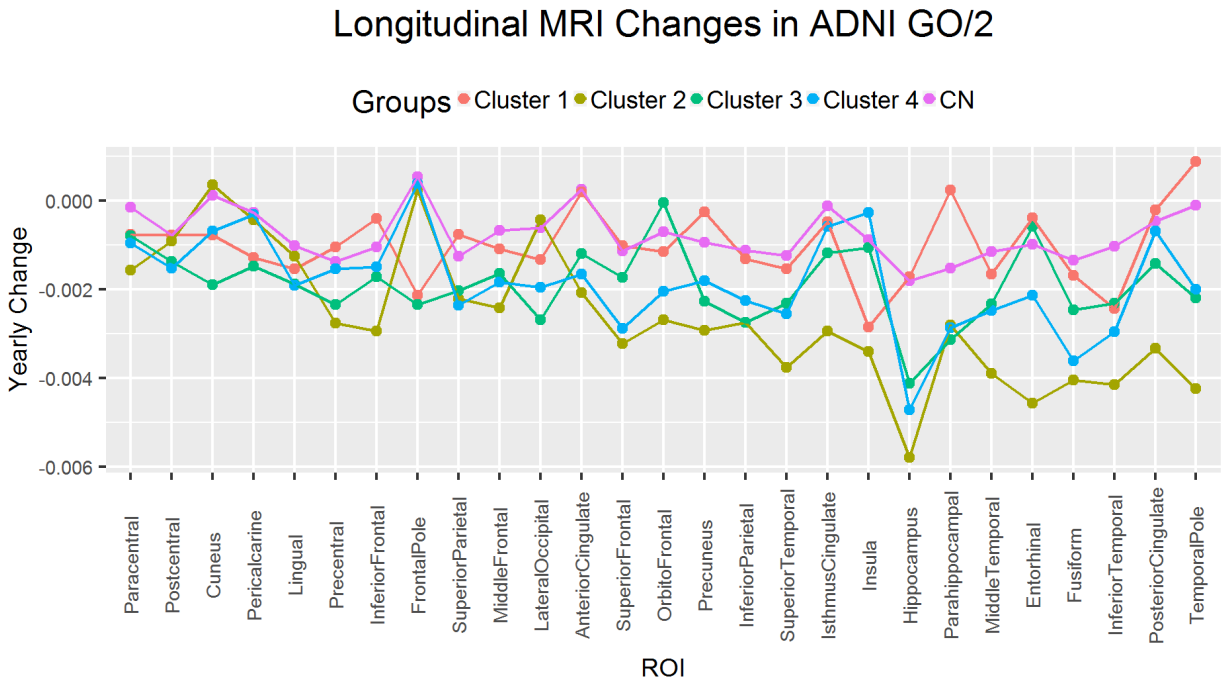
Supplementary Figure 6. Cognitive longitudinal changes based on aHV quartiles. Quartile 1 represents lowest volume, whereas Quartile 4 is the highest.



Supplementary Figure 7. Standardized yearly MRI changes observed in CN subjects and MCI subjects belonging to the four identified clusters. A total of 124 CN, 57 cluster 1, 44 cluster 2, 18 cluster 3 and 40 cluster 4 subjects were included in the analysis in ADNI-1 (A). 84 CN, 15 cluster 1, 17 cluster 2, 17 cluster 3 and 13 cluster 4 subjects were included in the analysis in ADNI-GO/2 (B).



(A)



(B)