**Supplementary Material**

**Supplementary Methods**

**1. Functional Connectivity Index of Regions of Interest:** **Regional-based functional connectivity is adopted in this study.**

 This study calculated the functional connectivity index (FCI) of three regions of interest (ROIs): bilateral hippocampus (HIPFCI), posterior cingulate cortex (PCCFCI), and fusiform gyrus (FUSFCI). Supplementary Figure 1 depicts the FCI calculation stream. First, the whole cerebral cortex was separated into 90 regions based on the Automated Anatomical Labeling (AAL) template, and the blood oxygen level dependent (BOLD) time series of each region was extracted using the AAL template mask from the preprocessed resting-state dataset [1]. Second, functional connectivity between each ROI and the other brain regions was calculated using the Pearson cross-correlation analysis. Thus, a vector consisting of 89 cross-correlation coefficient (CC) values for each ROI was obtained. Finally, each ROI’s FCI value—identified separately as HIPFCI, PCCFCI, and FUSFCI—was calculated by summating 89 CC values within each ROI’s vector and averaging them across each pair of bilateral ROIs to represent the functional connectivity strength.

**2. Gray Matter Index**

 For each individual, the gray matter index (GMI) [2] of each brain region (using the same AAL template) was calculated using SPM8 software (www.fil.ion.ucl.ac.uk/spm/software/spm8/). First, the anatomical image of each individual’s brain was normalized into the MNI space. Second, the gray matter of the whole brain was segmented and separated from white matter and cerebrospinal fluid (CSF) areas, and a threshold of 0.8 was used to exclude non-gray-matter areas. Third, each region’s GMI was determined by summing the gray matter concentration values of all voxels within the region and averaging across each pair of bilateral ROIs.

**3. Event-Based Probabilistic Model**

 Much of this section, describing the event-based probabilistic model, was adapted from [3]. Given that a set of N events, E1, E2, …, EN, is measured by N biomarkers, x1, x2, …, xN, respectively, the temporal operating sequence (TOS) of events, S = {s(1), s(2), …, s(N)}, is calculated by a permutation of the integers 1, …, N. For subjects j = 1, …, J, the dataset X could be regarded as X = {X1, X2, …, XJ}. Specifically, Xj represents the subject j data that is given by Xj = {x1j, x2j, …, xNj}, where xij is the ith biomarker measurement for subject j. This study determined the optimal TOS in a data-driven manner, based on the criteria that the optimal TOS, defined as the Soptimal, yielded the highest probability in measuring dataset X. That is, the p(X|S) value of the sequence Soptimal was calculated to be maximal among all of the possible sequences. To accomplish this objective, we first estimated the likelihood of measurement xij given that biomarker event Ei has or has not occurred. These likelihoods are labeled below:

$p(x\_{ij}\left|E\_{i})=likelihood of measurement x\_{ij} given that event E\_{i} has occurred \right. $ (1)

$p(x\_{ij}\left|¬E\_{i})=likelihood of measurement x\_{ij} given that event E\_{i} has not occurred \right.$ (2)

 We assumed that subject j is at stage k, although the authentic biomarker sequence and the subject’s stage were unavailable. This means that for subject j, events Es(1), Es(2), …, Es(k) already have occurred, and events Es(k+1), Es(k+2), …, Es(N) have not occurred. The likelihood of data Xj given the sequence S and the subject’s stage at k was obtained using the formula below:

$p\left(S,k\right)= \prod\_{i=1}^{k}p\left(E\_{s(i)}\right)\prod\_{i=k+1}^{N}p\left(¬E\_{s(i)}\right) $ (3)

Where $\prod\_{i=1}^{k}p\left(E\_{s(i)}\right)$ represents the overall likelihood of measurements given that corresponding events have already occurred, $\prod\_{i=k+1}^{N}p\left(¬E\_{s(i)}\right)$ is the overall likelihood of measurements given that events have not yet occurred. Then, we obtained the likelihood of data Xj in the condition of sequence S by summating the likelihood values of data Xj across all possible stages within sequence S, as shown in equation (4) below:

$ p\left(S\right)= \sum\_{k=0}^{N}p\left(k\right)p\left(S,k\right) $ (4)

Next, we combined the measurements of all subjects, j=1, …, J, assuming that the intersubject relationship is independent:

$$p\left(S\right)=\prod\_{j=1}^{J}p\left(S\right)=\prod\_{j=1}^{J}\sum\_{k=0}^{N}p\left(k\right)p\left(S,k\right)$$

$ =\prod\_{j=1}^{J}\sum\_{k=0}^{N}p\left(k\right)\left[\prod\_{i=1}^{k}p\left(E\_{s(i)}\right) \prod\_{i=k+1}^{N}p\left(¬E\_{s(i)}\right)\right] $ (5)

In theory, the above analysis needs to be repeated for each possible sequence to determine the sequence (Soptimal) with the maximal value of $p\left(S\right)$. However, such a computation strategy is extremely time-consuming; total calculation times in this study would be 2.7942e+009, given that there are 10 biomarker events, 11 possible stages (including stage 0), and 70 subjects (cognitive normal [CN] and Alzheimer’s disease [AD] groups only) involved. Therefore, we employed a greedy algorithm to improve processing efficiency.

**4. Event Occurrence and Nonoccurrence Distribution Modeling**

 We used a mixture model of two Gaussian distributions to fit the event data from the CN and AD groups, based on the assumption that an event occurring and an event not occurring are estimated by a mixed distribution of normal and abnormal groups. Although the distribution of biomarker values is not pure Gaussian, a previous study has shown the Gaussian mixture model had similar results with a uniform model for the AD cohort [3]. The fitted Gaussian distributions separated the data into two groups, i.e., abnormal (event occurred) and normal (event did not occur), similar to the approach by Young et al. [4]. Notably, we modified Young et al.’s approach by applying a k-means clustering algorithm to separate the whole distribution into two clusters before applying the Gaussian mixture model fitting. This modified modeling method led to high consistency in the obtained model.

**5. Self-Growing Greedy Algorithm**

 The greedy algorithm explores the globally optimal solution by making the locally optimal choice at each stage, in a greedy heuristic manner. The greedy Markov chain Monte Carlo (MCMC) algorithm is a useful approach to find globally optimal results.

 The amount of time such an analysis would take is unpredictable due to the randomized initial sequence, and it may be quite long due to the inevitable searching loop. Therefore, we developed a new greedy algorithm to address this deficiency. Specifically, we started with a set of all possible initial root sequences, each of which consisted of two randomly selected events from the 10 biomarker events total. Second, for each initial sequence S, we generated the children of S by inserting a randomly selected event from the remaining events. Third, we selected the children sequence with the maximal $p\left(S\right)$ value; this replaced the initial sequence. Then, we entered another randomly selected event into the sequence and repeated the second and the third steps until no events were left. Thus, we generated whole sequences for each root sequence. Ultimately, we determined the sequence with the maximum $p\left(S\right)$ value as the final optimal sequence, Soptimal. We repeated this greedy algorithm 100 times to ensure the Soptimal had a high reliability. This study used 45 CN and 25 AD subjects to determine the Soptimal. Note that the early mild cognitive impairment (EMCI) and late mild cognitive impairment (LMCI) subjects were not used to train the Soptimal.

 The Soptimal reflects the order in which sequential pathophysiological events occurred and provides a numeric score to measure disease progression from one stage to the next.

**6. Individual Staging Based on the Obtained Sequence**

 Using the following equation to determine each subject’s AD risk stage, we calculated the likelihood value of *k* at each possible stage in the sequence and defined the AD stage as that at which *k* had the highest likelihood value at the Soptimal:

$argmax\_{k}P\left(k\right)= argmax\_{k}(\prod\_{i=1}^{k}p\left(E\_{S^{optimal}(i)}\right) \prod\_{i=k+1}^{N}p\left(¬E\_{S^{optimal}(i)}\right)) $ (6)

In equation 6, implications of $\prod\_{i=1}^{k}p\left(E\_{S^{optimal}(i)}\right)$ and $\prod\_{i=k+1}^{N}p\left(¬E\_{S^{optimal}(i)}\right)$ refer to those in equation 3, except that the optimal sequence, Soptimal, is obtained.

**7. Statistical Analysis**

 We employed multiple linear regression models to estimate the relationship between the characterizing Alzheimer’s disease risk events (CARE) index scores and Auditory Verbal Learning Test-30-min delayed recall (AVLT30min) scores across the four groups, as follows:

$$CARE Index\left(s\right)=$$

$ b0+b1·I\left(s,EMCI\right)+b2·I\left(s,LMCI\right)+b3·I\left(s,AD\right)+ b4·AVLT30min\left(s\right)+b5·AVLT30min\left(s\right)·I\left(s,EMCI\right)+b6·AVLT30min\left(s\right)·I\left(s,LMCI\right)+b7·AVLT30min\left(s\right)·I\left(s,AD\right)+ε(s)$ (7)

where *s* is the subject variable and the group indicator variable *I(s,EMCI)* equals1 if subject *s* is within the EMCI group, 0 otherwise, etc. This model allows the simultaneous fitting of four regression lines to the data and corrects for errors due to the multiple group effect. We also employed a nonlinear logistic model to fit to all of the data:

$CARE Index\left(s\right)=b0+A·\left(e^{b2+b3(b4-AVLT30min\left(s\right))}\right)/\left(1+e^{b2+b3(b4-AVLT30min\left(s\right))}\right)$ (8)

The unknown parameters (b0, A, b2, b3, b4) in the above model were estimated using a nonlinear least squares algorithm.

**8. Intrasubject Repeatability of CARE Index measurement**

 We compared intrasubject CARE index consistency by repeating individual-subject measurements using the short-term longitudinal (six-month) Alzheimer's Disease Neuroimaging Initiative 2 datasets (ADNI 2) [5]. We limited our analysis to a six-month period to ensure a sufficient number of study subjects and minimize the time effect on disease progression. We chose the subjects with a second MRI (magnetic resonance imaging) scan visit within 75 and 180 days from their baseline visits. This resulted in 33 subjects from the ADNI 2 dataset (with resting-state functional connectivity MRI subset data). Because CSF samples for amyloid-β (Aβ) and phosphorylated tau (p-tau) measurements were not available at the second visit, the Aβ and p-tau values for the second visit were interpolated between the first and second CSF samples that were taken two years apart.

**REFERENCES**

[1] Tzourio-Mazoyer N1, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B, Joliot M (2002) Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* **15**, 273-289.

[2] Xie W, Song C, Young NL, Sperling AS, Xu F, Sridharan R, Conway AE, Garcia BA, Plath K, Clark AT, Grunstein M (2009) Histone h3 lysine 56 acetylation is linked to the core transcriptional network in human embryonic stem cells. *Mol Cell* **33**, 417-427.

[3] Fonteijn HM, Modat M, Clarkson MJ, Barnes J, Lehmann M, Hobbs NZ, Scahill RI, Tabrizi SJ, Ourselin S, Fox NC, Alexander DC (2012) An event-based model for disease progression and its application in familial Alzheimer's disease and Huntington's disease. *Neuroimage* **60**, 1880-1889.

[4] Young AL, Oxtoby NP, Daga P, Cash DM, Fox NC, Ourselin S, Schott JM, Alexander DC; Alzheimer’s Disease Neuroimaging Initiative (2014) A data-driven model of biomarker changes in sporadic Alzheimer's disease. *Brain* **137**, 2564-2577.

[5] Rezaee K, Haddadnia J (2013) Designing an algorithm for cancerous tissue segmentation using adaptive K-means cluttering and discrete wavelet transform. *J Biomed Phys Eng* **3**, 93-104.

Supplementary Figure 1:

