

1 Quantitative assessment on the severity degree of Alzheimer dementia
2 by algebraic analysis on cortical thickness profiles of human brains
3

4 Sangyeol Kim^{1, 2, \$, #}, Seongjun Park^{3, \$}, Iksoo Chang^{1, 2, 4, #}

5 and the Alzheimer's Disease Neuroimaging Initiative*

6

7 ¹ Department of Brain and Cognitive Sciences, Daegu Gyeongbuk Institute of Science
8 and Technology, Daegu, Korea

9 ² Brain and Cognitive Sciences Research Institute, Daegu Gyeongbuk Institute of
10 Science and Technology, Daegu, Korea

11 ³ Department of Emerging Materials Science, Daegu Gyeongbuk Institute of Science
12 and Technology, Daegu, Korea

13 ⁴ Supercomputing Bigdata Center, Daegu Gyeongbuk Institute of Science and
14 Technology, Daegu, Korea

15

16 \$ Contributed equally to this work.

17 # Correspondence: Sangyeol Kim (sykim@dgist.ac.kr) and

18 Iksoo Chang (iksoochang@dgist.ac.kr)

19

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

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

1 Abstract

2 Alzheimer disease(AD) affects profoundly the quality of human life. Quantifying the
3 severity degree of Alzheimer dementia for an individual person is critical for the early
4 diagnose and prescription for delaying the further dementia progression. However, the
5 quantitative diagnose for human subjects of the mild cognitively impairment(MCI) or
6 AD with the different degree of dementia severity is a difficult task due to both the very
7 broad distribution of dementia severities and the lack of good quantitative determinant
8 to assess it. Here, based on cortical thickness profiles of human brains measured by
9 magnetic resonance experiment, a new algebraic approach is presented for the
10 personalized quantification of the severity degree of AD, ranging from 0 for the basin
11 of cognitively normal state to 1 for AD state. Now one can unfold the broad distribution
12 of dementia severities and corresponding cortex regions of human brains with MCI or
13 AD by the different severity degree.

14

1 **Introduction**

2 Alzheimer's disease (AD) is one of the most well-known neurodegenerative diseases,
3 and it profoundly affects the quality of human life. Various causes of AD are known,
4 and clinical treatments for preventing or delaying the progression of AD are currently
5 practiced. The precise diagnosis of AD requires not only the systematic identification
6 of cohorts that classify the different stages of brain progression toward AD but also the
7 estimation of the severity degree of AD for a given individual (Galvin et al., 2012,
8 Solomon and Soininen, 2015, Raj et al., 2012). The symptoms of AD appear in various
9 forms in the human body, behavior, and cognition, yet the direct anatomical evidences
10 appear in the structural change within the brain cortex (Hojjati et al., 2019, Kim et al.,
11 2020, Qiu et al., 2020, Tetreault et al., 2020). Among these evidences is the
12 degradation of the cortical thickness of the human brain, which is one of the imprints
13 of AD. Such physical change can be monitored through the neuro-image analysis, for
14 example the magnetic resonance image (MRI) analysis of the brain (Hartikainen et al.,
15 2012, Im et al., 2008, Kim et al., 2014, Lebedev et al., 2013, Paternico et al., 2016,
16 Querbes et al., 2009).

17 The anatomical degradation of the cortical thickness becomes more pronounced as
18 the degree of dementia becomes greater (Im et al., 2008, Querbes et al., 2009).
19 Therefore, there is a rationale for taking advantage of the degree of cortical
20 degradation for determining the dementia cohort and estimating the severity of
21 dementia progression. Given the big data information of cortical thickness for a single
22 subject, however, we are faced with several obstacles to overcome in accomplishing
23 abovementioned tasks. First, we noted that the person-to-person fluctuations in
24 cortical thickness of an individual may overwhelm the degradation in cortical thickness.
25 In clinical cases, we frequently observed that the average cortical thickness of some

1 cognitively normal people is thinner than that of people with AD, which appears to
2 contrast conventional view. The second obstacle is the ambiguity in what we should
3 do if two different cohorts have a difference in the cortical thickness in brain regions
4 that have little to do with AD. In principle, we should construct some determinants for
5 distinguishing a cognitively normal person from a person with AD based on cortical
6 thickness, but in practice, we are confronted with the differences in cortical thickness
7 in regions of the cortex that are unrelated to the pathogenesis of AD. We noted that
8 the abundant existence of such unrelated regions is an intrinsic source that increases
9 the uncertainty of the dementia determinants and hinders the appropriate construction
10 of good classifier and predictor for AD.

11 In this study, instead of dealing with all 327,684 vertices point on the whole cortex of
12 a human brain, we demonstrated that the consideration of a few hundred essential
13 vertices were enough for distinguishing CN, MCI, AD cohorts each other. With cortical
14 thickness data at these essential vertices, we defined the statistical score matrix and
15 the covariance correlation matrix between human subjects as a set of classifier and
16 predictor for AD. Unlike the conventional view that the degradation of the cortical
17 thickness of human brain was responsible for AD, the singular valued decomposition
18 analysis of the statistical score matrix developed in this study implicated that the
19 simultaneous consideration of both thinner and thicker cortical regions together
20 compared to those of CN are important and necessary for the precise diagnose of the
21 severity of AD.

22

23

1 **Materials and methods**

2 **Preparation of cortical thickness data from MRI of 1522 human brain 3 images from ADNI**

4 Data used in the preparation of this article were obtained from the Alzheimer's Disease
5 Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched
6 in 2003 as a public-private partnership, led by principal investigator Michael W. Weiner,
7 MD. The primary goal of the ADNI has been to test whether serial MRI, PET, other
8 biological markers, and clinical and neuropsychological assessment can be combined
9 to measure the progression of MCI and early Alzheimer's disease. For up-to-date
10 information, see www.adni-info.org.

11 We selected 274 individuals (human subjects) who were identified as CN, 265
12 individuals with MCI, 125 individuals with AD from the ADNI-2 study of ADNI, and 97
13 with MCI from the ADNI-GO study of ADNI. A human brain image-data set of 1522 MR
14 images from a total of 761 subjects was constructed, for each of which both 1.2-mm
15 sagittal Magnetization Prepared Rapid Gradient Echo (MPRAGE) and
16 MPRAGE_SENSE2 images were taken separately.

17

18 **Partition 1516 MR images of human brains into four groups and determine 19 the essential region-of-interest vertices for each group**

20 We performed the FreeSurfer analysis to obtain the cortical thickness data at 327,684
21 vertices on the cortex of a human brain (Dale et al., 1999, Fischl et al., 1999). The
22 cortical thickness at each vertex ranges from 0 mm to 5 mm. After eliminating those
23 vertices at which cortical thickness information was missing for any one of the 1522
24 MR images of human brains in the ADNI data set, 276,825 common vertices whose

1 cortical thickness values are available for all 1522 MR images were selected for our
2 study. The average cortical thickness over 276,825 vertices for each brain images was
3 evaluated, and we divided 1516 values of average thickness into four groups (A-D) of
4 different windows of average thickness except 6 values of that run out-of-bounds.
5 Demographic characteristics of the average cortical thickness of the four groups are
6 listed in Table 1.

Table 1. Demographic Characteristics of the groups

Group A	CN	MCI	AD
N*	136	155	13
female (%)	60.3	53.5	30.8
Age, Mean (SD)	72.5 (5.3)	72.1 (5.1)	78.1 (5.0)
$\langle t \rangle^{\dagger}$, Mean (SD)	2.4580 (0.0477)	2.4549 (0.0415)	2.4676 (0.0539)
Group B	CN	MCI	AD
N	212	262	67
female (%)	62.7	43.5	41.8
Age, Mean (SD)	73.0 (5.2)	73.7 (5.3)	76.7 (6.3)
$\langle t \rangle$, Mean (SD)	2.3516 (0.0293)	2.3429 (0.0266)	2.3491 (0.0269)
Group C	CN	MCI	AD
N	159	209	84
female (%)	46.5	34.0	52.4
Age, Mean (SD)	74.0 (5.7)	75.3 (5.8)	76.9 (5.2)
$\langle t \rangle$, Mean (SD)	2.2566 (0.0290)	2.2573 (0.0278)	2.2455 (0.0298)
Group D	CN	MCI	AD
N	40	96	83
female (%)	22.5	35.4	26.5
Age, Mean (SD)	78.1 (6.5)	76.2 (5.7)	77.4 (6.6)
$\langle t \rangle$, Mean (SD)	2.1514 (0.0389)	2.1448 (0.0433)	2.1373 (0.0520)

* Number of MRI images; † Average cortical thickness over the 276,825 vertices. AD,

Alzheimer's disease; CN, cognitively normal; MCI, mild cognitive impairment.

1 In order to assign subjects from each CN, MCI, and AD cohort into one of the four
2 groups (A-D) of average cortical thickness, we employed the Z score criteria in
3 selecting the region-of-interest (ROI) vertices and the essential ROI vertices on the
4 cortex at which the distribution of cortical thickness of the CN cohort is distinguished
5 from that of the AD cohort within each group of average cortical thickness. A similar
6 procedure is repeated for distinguishing the CN cohort from the MCI cohort and also
7 the MCI cohort from the AD cohort:

$$8 \quad Z_p^{\text{CN-MCI}} = \frac{\langle t_{p,h \in \text{CN}} \rangle - \langle t_{p,h \in \text{MCI}} \rangle}{\sqrt{\frac{\sigma_{p,h \in \text{CN}}^2}{n_{p,h \in \text{CN}}} + \frac{\sigma_{p,h \in \text{MCI}}^2}{n_{p,h \in \text{MCI}}}}}, \quad Z_p^{\text{CN-AD}} = \frac{\langle t_{p,h \in \text{CN}} \rangle - \langle t_{p,h \in \text{AD}} \rangle}{\sqrt{\frac{\sigma_{p,h \in \text{CN}}^2}{n_{p,h \in \text{CN}}} + \frac{\sigma_{p,h \in \text{AD}}^2}{n_{p,h \in \text{AD}}}}}, \quad Z_p^{\text{MCI-AD}} = \frac{\langle t_{p,h \in \text{MCI}} \rangle - \langle t_{p,h \in \text{AD}} \rangle}{\sqrt{\frac{\sigma_{p,h \in \text{MCI}}^2}{n_{p,h \in \text{MCI}}} + \frac{\sigma_{p,h \in \text{AD}}^2}{n_{p,h \in \text{AD}}}}}. \quad (1)$$

9 Here, $\langle t_{p,h \in k} \rangle$ is the average cortical thickness at a vertex point p averaged over the
10 subject h who belongs to the k (one of CN, MCI, AD) cohort, and $\sigma_{p,h \in k}$ is its standard
11 deviation, and $n_{p,h \in k}$ is the number of MR images belonging to the k cohort. The
12 positive (negative) value of $Z_p^{\text{CN-AD}}$, for example, indicates that the distribution curve
13 of the average cortical thickness of the CN cohort is right (or left)-shifted compared to
14 that of AD cohort. And the bigger the absolute value of the Z score, the better
15 distinguished the distribution curves of average cortical thickness of the cohorts. In
16 this study, we identified ROI vertices satisfying the absolute value of the Z score larger
17 than 1.5, and essential ROI vertices satisfying much higher cut-off Z scores (Table S1).

18

19 **Construction of a statistical score matrix for classifying subjects into one
20 of CN, MCI, AD cohorts**

21 Within each group of average cortical thickness, we constructed the statistical score

1 matrix for determining a subject's cohort as either CN, MCI, or AD (Yu et al., 2011).

2 First of all, $t_{p,h \in k}$ was transformed into the probability distribution matrix $P_{p,m}^{(k)}$, which

3 is a probability that the cortical thickness $t_{p,h \in k}$ at a vertex point p of the subject in k

4 cohort is between $(m - 1)\Delta$ and $m\Delta$:

5
$$P_{p,m}^{(k)} = \frac{\sum_{h \in k} \Theta(t - (m-1)\Delta) \cdot \Theta(m\Delta - t) \cdot \delta(t - t_{p,h})}{\sum_{h \in k} 1}, \quad m = 1, 2, \dots, 30, \quad k = \text{CN, MCI, AD}. \quad (2)$$

6 Here, $\Delta = 0.2$ mm, and cortical thickness index m runs from 1 to 30; this covers the

7 cortical thickness from 0 mm to 6 mm. $\delta(x)$ is a Dirac delta function, and $\Theta(x)$ is a

8 step function where $\Theta(x < 0) = 0; \Theta(x \geq 0) = 1$. Then, we defined the statistical score

9 matrix $S_{p,m}^{(k)}$ from $P_{p,m}^{(k)}$ in the following way:

10
$$S_{p,m}^{(k)} = -\ln \left[\frac{Q_{p,m}^{(k)}}{Q_p^{(k)}} \right], \quad Q_{p,m}^{(k)} = \frac{P_{p,m}^{(k)}}{\sum_m P_{p,m}^{(k)}}, \quad Q_p^{(k)} = \frac{\sum_m P_{p,m}^{(k)}}{\sum_p \sum_m P_{p,m}^{(k)}}. \quad (3)$$

11 Since $\sum_m P_{p,m}^{(k)} = 1$, $S_{p,m}^{(k)} = -\ln P_{p,m}^{(k)} - \ln \sum_p 1$ and the second term are constants. The

12 value of the statistical score matrix $S_{p,m}^{(k)}$ varies depending on the cohort k ; the smaller

13 $S_{p,m}^{(k)}$ is, the larger $P_{p,m}^{(k)}$ is.

14 With this statistical score matrix $S_{p,m}^{(k)}$, we employed a strategy for determining to which

15 one of k cohorts a given subject would belong. First, we evaluated the averaged

16 cortical thickness of a given subject over 276,825 vertices. Second, we assigned this

17 subject to one of four groups (A-D) of average cortical thickness. Third, based on the

18 preselected essential ROI vertices p for the assigned group, we determined the

19 cortical thickness index $m'(p)$ at which the cortical thickness at an essential ROI vertex

20 p is between $(m - 1)\Delta$ and $m\Delta$. Then, for each k cohort, the total score $S'_{(k)}$ was

21 calculated by summing up $S_{p,m'(p)}^{(k)}$ over the preselected essential ROI vertices p for

22 the assigned group of the average cortical thickness, $S'_{(k)} = \sum_p S_{p,m'(p)}^{(k)}$. Lastly, to

1 which k cohort a given subject would belong was decided by a cohort which gives the
2 minimum score out of $S'_{(CN)}$, $S'_{(MCI)}$, $S'_{(AD)}$.

3 We, however, noted that the accuracy of both $P_{p,m}^{(AD)}$ and $S_{p,m}^{(AD)}$ may become
4 unsatisfactory if the number of people in the AD cohort was less than that of the CN
5 cohort and the MCI cohort (Table 1). In order to overcome the unsatisfactory nature of
6 both $P_{p,m}^{(AD)}$ and $S_{p,m}^{(AD)}$, we employed the method of Kernel Density Estimation (KDE);
7 namely, a Dirac delta function $\delta(t - t_{p,h})$ in the definition of the probability distribution
8 matrix $P_{p,m}^{(k)}$, is replaced by a kernel function $f(t - t_{p,h})$:

$$9 \quad f(t - t_{p,h}) = \sum_{l=-4}^4 a_l \delta(t - (t_{p,h} + l\Delta)), \quad (4)$$

$$10 \quad a_l = \text{erf}(0.25 + 0.5|l|) + \frac{a_0}{2} - \sum_{l'=0}^{|l|-1} a_{l'} . \quad (5)$$

11 Here, the relative ratio among coefficients a_l is $a_0 : a_{\pm 1} : a_{\pm 2} : a_{\pm 3} : a_{\pm 4} = 56 : 43 : 21 : 7 : 1$.
12 The kernel function $f(t - t_{p,h})$ satisfies $\int_{-\infty}^{\infty} f(t - t_{p,h}) dt \approx 1$ and the standard
13 deviation $\sigma_f \approx 1.435$. Upon subjecting KDE, $P_{p,m}^{(k)}$, becomes

$$14 \quad P_{p,m}^{(k)} = \frac{\sum_{h \in k} \Theta(t - (m-1)\Delta) \cdot \Theta(m\Delta - t) \cdot f(t - t_{p,h})}{\sum_{h \in k} 1}, \quad m = 1, 2, \dots, 30, \quad k = \text{CN, MCI, AD}. \quad (6)$$

15 In this study, we constructed the statistical score matrix on which KDE was employed
16 and used it for determining to which k cohort a given subject would belong.

17

18 **Construction of a covariance correlation matrix and a predictor for the
19 severity degree of AD**

20 Within each group of the average cortical thickness, the severity degree of AD for a
21 given subject is estimated by the following strategy. First of all, we transformed the
22 cortical thickness matrix $t_{p,h}$ at essential ROI vertices p for a subject h into the

1 normalized matrix $t'_{p,h}$ such that

2
$$t'_{p,h} = \frac{t_{p,h} - \langle t_{p,h} \rangle_h}{\sigma_p}, \quad \langle t_{p,h} \rangle_h = \frac{\sum_h t_{p,h}}{\sum_h 1}, \quad \sigma_p^2 = \langle t_{p,h}^2 \rangle_h - \langle t_{p,h} \rangle_h^2. \quad (7)$$

3 Here, the product of $t'_{p,j}$ by its transpose $t'_{p,i}^T$ results in the square matrix $t''_{ij} =$

4 $t'_{p,i}^T \cdot t'_{p,j}$, and then its normalized matrix (called by a covariance correlation matrix)

5 C_{ij} is defined by $C_{ij} = t''_{ij} / \max\{t''_{ij}\}$, where $\max\{t''_{ij}\}$ is the maximum value of

6 elements in the square matrix t''_{ij} . The larger the value of C_{ij} , the higher the

7 covariance correlation between a subject i and a subject j in their profile of the cortical

8 thickness at essential ROI vertices. Based on this covariance correlation matrix, we

9 defined the severity degree AD for a given subject i by

10
$$SD_i = (\overline{C_i^{AD-CN}} - \langle C_i^{AD-CN} \rangle_{i \in CN}) / (\langle C_i^{AD-CN} \rangle_{i \in AD} - \langle C_i^{AD-CN} \rangle_{i \in CN}), \quad (8)$$

11 where $\overline{C_i^{AD-CN}} = \langle C_{ij} \rangle_{j \in AD} - \langle C_{ij} \rangle_{j \in CN}$. The severity degree of AD ranged from 0 for the

12 basin of CN state to 1 for the basin of AD state. Rank-ordering this degree in ascending

13 order illustrates that a subject i with the larger (or smaller) value of the severity degree

14 is more prone to AD (CN) state.

15

16

1 Results

2 Identification of essential ROI vertices at which the distributions of 3 cortical thickness of CN, MCI, AD subjects are distinguishable

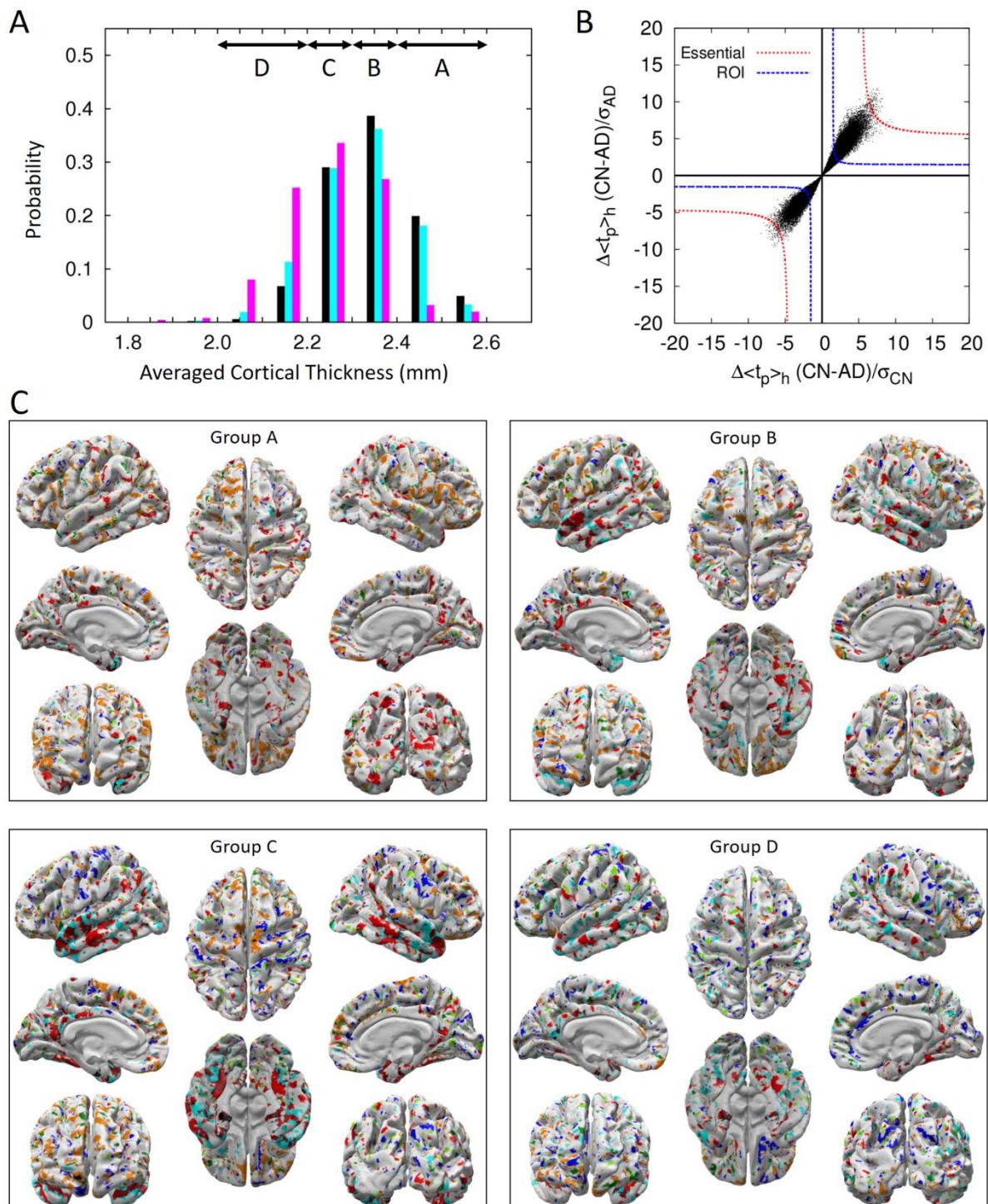
4 Although the averaged cortical thickness of subjects with AD is generally known to be
5 thinner than that of CN or MCI subjects, the distribution curves of averaged cortical
6 thickness for the cohorts are not well distinguishable except near both ends of the
7 distribution curves as demonstrated in Figure 1A. This illustrates that a subject can be
8 CN even though the averaged cortical thickness is thinner than that of a subject with
9 AD, and vice versa. Also, we found that many subjects identified as CN, MCI, or AD
10 have a similar averaged cortical thickness. This is due to the fact that the average
11 cortical thickness for a subject was calculated over all 276,825 vertices, and the
12 cortical thicknesses at most vertices are similar for all subjects, which prohibits us from
13 successfully clustering 1552 human brain images into the image of CN, MCI, AD
14 cohorts. Therefore, instead of resorting on the cortical thickness of all 276,825 vertices,
15 we need to select the ROI vertices at which the cortical thickness values are
16 distinguishable from each other among CN, MCI, and AD. For a fair selection of such
17 ROI vertices, we divided the range of the averaged cortical thickness of subjects into
18 four (A-D) different regions (for the detailed method, see in the section 2.2).

19 Figure 1B illustrates how we identified ROI and essential ROI vertices. The x-axis
20 $(\langle t_{p,h \in CN} \rangle - \langle t_{p,h \in AD} \rangle) / \sqrt{\sigma_{p,h \in CN}^2 / n_{p,h \in CN}}$ represents the degree of separation between
21 the distribution curves of cortical thickness for CN subjects and AD subjects at a vertex
22 p normalized by the dispersion of the cortical thickness of CN subjects, which is
23 quantified by the value of its Z score. The y-axis

1 $(\langle t_{p,h \in CN} \rangle - \langle t_{p,h \in AD} \rangle) / \sqrt{\sigma_{p,h \in AD}^2 / n_{p,h \in AD}}$ represents values normalized by the
2 dispersion of the cortical thickness of individuals with AD. Therefore, the x-values (y-
3 values) at a vertex point p represent the degree by which the distribution of cortical
4 thickness at this point p of CN (AD) subjects is distinguished from averaged cortical
5 thickness of subjects with AD (CN). It means the larger the value of $(Z_p^{CN-AD})^2 =$
6 $[x^{-2} + y^{-2}]^{-1}$ is, the two distribution curves are more distinguished each other (for the
7 illustration, see Figure S1). The ROI cut-off line is defined by points satisfying
8 $|Z_p^{CN-AD}| = 1.5$, and the distribution of cortical thickness of CN subjects and individuals
9 with AD is clearly distinguished at those points satisfying $|Z_p^{CN-AD}| > 1.5$ (outside of
10 the ROI cut-off line).

11 Figure 1C shows ROI vertices for each of the cortical thickness groups (A-D) by
12 colored points on the white cortex, at which the thickness of the cortex is either thicker
13 or thinner particularly for one cohort compared with that of two other cohorts. These
14 ROI vertices are widely distributed on the cortex, and their locations are not fixed but
15 vary depending on the groups A to D. We uncovered, however, that the medial
16 temporal lobe, known to be very important for memory formation, is always indicated
17 by a red or dark red color irrespective of the groups A to D (Figure S2). This implies
18 that the cortical thickness values of the medial temporal lobe for subjects with AD are
19 characteristically thinner than those of CN subjects or individuals with MCI (red color),
20 and this decrease occurs in the following descending order: CN-MCI- AD (dark red
21 color). The medial temporal lobe is the region where the cortical thickness gradually
22 decreases as dementia progresses and therefore is the critical region necessary for
23 determining the dementia cohort and the severity degree of AD. We also noted that
24 the cortical thickness at the orange-colored region for subjects with AD is thicker than

- 1 that for CN subjects or those with MCI. This has nothing to do with the damage in the
- 2 cortex but contributes to the increase in the accuracy for predicting the dementia
- 3 cohort since it could provide better distinguishability of subjects with AD from CN
- 4 subjects and those with MCI.

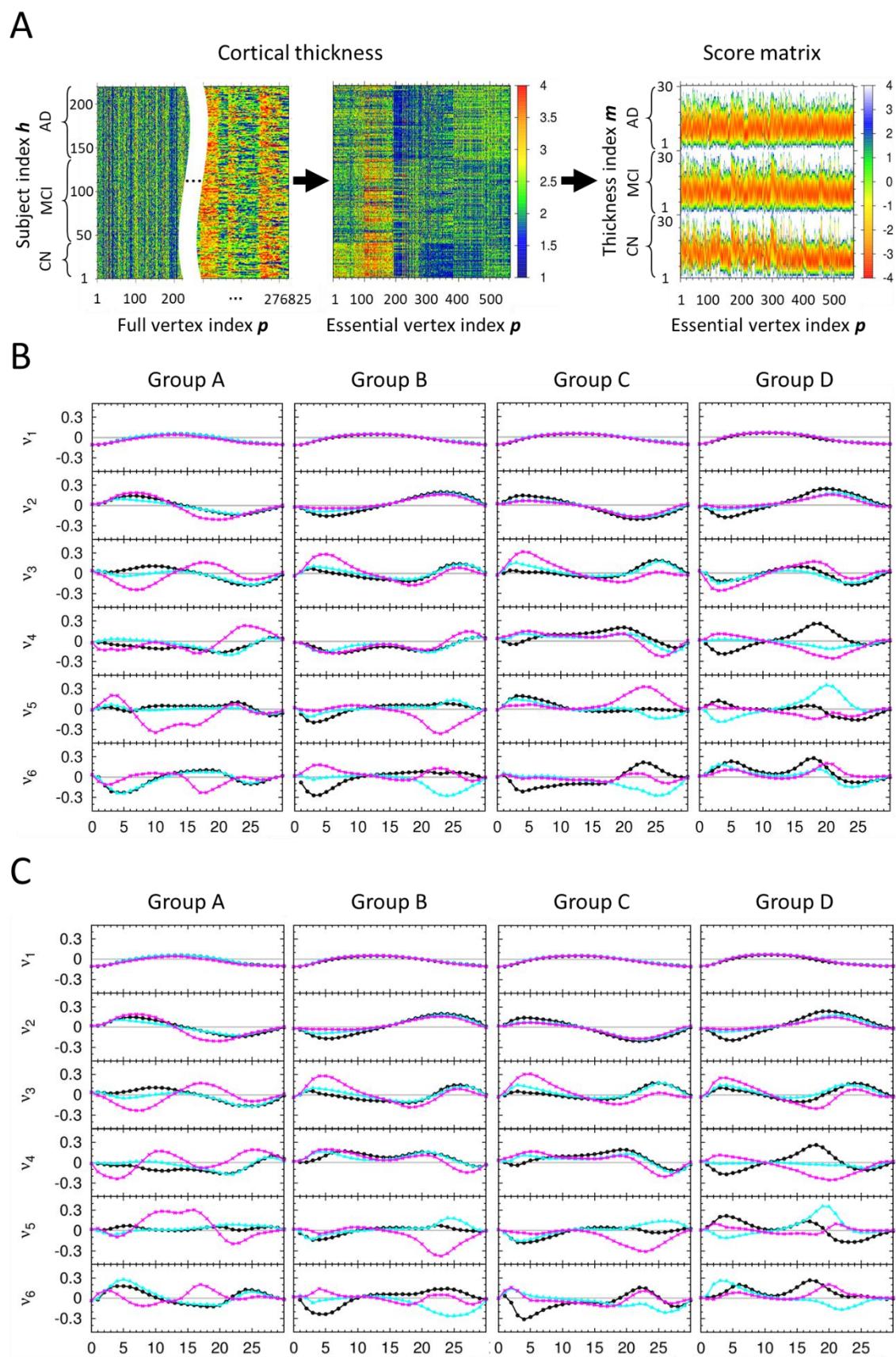


1 **Figure 1. The classification of 1516 brain images into four groups by the average**
2 **cortical thickness of subjects from the ADNI.** (A) The distribution of average cortical
3 thickness of subjects in the CN (black), MCI (cyan), and AD (magenta) cohorts. Above
4 arrows point to the range of average cortical thickness. Subject number, sex, and age for
5 each group are listed in Table 1. (B) For the cortical thickness group D, the degree of
6 separation of the distribution curve of average cortical thickness between CN subjects and
7 AD subjects is presented in the form of black points. The closer to the origin point (0, 0) the
8 degree of separation of two distribution curves of average cortical thickness is, the less
9 distinguishable they are (Figure S1). Black points residing outside of the blue-dashed line
10 ($Z = \pm 1.5$) are ROI vertices, and black points residing outside of red-dashed line (Z values
11 are listed in Table S1) are essential ROI vertices. (C) For each group of average cortical
12 thickness, ROI vertices at which the thickness of the cortex for CN subjects is thicker
13 (thinner) than that of the other subjects with MCI or AD is represented by cyan (blue) color.
14 As a similar procedure, ROI vertices for MCI subjects is thicker (thinner) than the other
15 cohorts is represented by green (dark green) color. Also, ROI vertices for AD subjects is
16 thicker (thinner) than the other cohorts is represented by orange (red) color. Especially, the
17 ROI vertices at which the cortical thickness decreases in the descending order of CN-MCI-
18 AD is represented by dark red. And the essential ROI vertices are represented by a black
19 color. The ROI vertices commonly found from more than three groups of average cortical
20 thickness are presented in Figure S2.

21

22 **Character of statistical score matrix and classification of subject's cohort**
23 The section 2.3 described the detailed procedure of constructing the statistical score
24 matrix for determining a subject's cohort within each group of the average cortical
25 thickness (Figure 2A). In order to judge how well the statistical score matrix
26 distinguishes CN, MCI, and AD cohorts from each other before we predict the cohort

1 of a new subject, we performed the singular value decomposition (SVD) analysis on
2 the combined statistical score matrix $S^{(All)}$ which consists of matrix elements of $S^{(CN)}$,
3 $S^{(MCI)}$, and $S^{(AD)}$. We used the SVD character of a matrix that a given matrix can be
4 reconstructed as a linear combination of the products of two singular vectors by one
5 singular matrix. Since the reconstructed matrix from the few highest modes of singular
6 values contains the predominant character of an original given matrix, one expects
7 that the differences among the cohorts should appear in singular vectors of different
8 cohorts. Otherwise, the statistical score matrix $S^{(All)}$ is not reliable nor does it contain
9 the characteristic ingredient of different cohorts. Figure 2B, 2C and S3 show the
10 highest six singular vectors corresponding to the six largest singular values from SVD
11 analysis of the statistical score matrices for each group of the average cortical
12 thickness A to D. Here, it demonstrates that elements in the singular vectors v_1 to v_3
13 for CN, MCI, and AD follow qualitatively a similar trend, meaning that these compose
14 the fundamental default modes, whereas those in v_4 to v_5 follow a different trend and
15 are distinguished each other.



1

2

Figure 2. The character of score matrix for each group of average cortical thickness.

1 (A) For a given group of average cortical thickness, three kinds of heat maps illustrate the
2 process starting from the cortical thickness matrix at all 276,825 vertices to that at only 564
3 essential ROI vertices, and then construction of the score matrix. The dimension in the x-
4 axis of the cortical thickness matrix at all 276,825 is too large to draw, we placed the blank
5 in the middle to abbreviate the large dimension of the x-axis. (B) The results of singular
6 value decomposition analysis on score matrices, which are composed of 547 CN, 722 MCI,
7 247 AD human brain images and used for self-recognition test. For each group of average
8 cortical thickness, six singular vectors corresponding to the six largest singular values are
9 presented. Here, x-axis is m value defined in the section 2.3, and y-axis is an arbitrary unit
10 for the singular vectors. For each graph, the singular vector components for CN, MCI, and
11 AD subjects are plotted by black, cyan, magenta colors, respectively. (C) The results of
12 singular value decomposition analysis on score matrices, which are composed of 363 CN,
13 480 MCI, 163 AD human brain images as a training set and used for the first iteration of
14 the stratified 3-fold cross validation test. The other results of that used for the second and
15 third iterations of the stratified 3-fold cross validation test are presented in Figure S3.

16

17 Out of 547 CN, 722 MCI, and 247 AD human brain images predetermined clinically
18 and provided from the ADNI data set, we performed the self-recognition test and the
19 stratified 3-fold cross validation test using the 1516 human brain images for each
20 group on the average cortical thickness A to D (Table 2). For the first (second; third)
21 iteration of stratified 3-fold cross validation test, 1006 (1011; 1015) human brain
22 images were used as the training set for learning the statistical score matrix and 510
23 (505; 501) human brain images were used as an independent validation set. The new
24 method presented in this study recognized and predicted the subjects with AD in the
25 cohort with more than 91% (self-recognition test) and 82% (stratified 3-fold cross
26 validation test) accuracy, respectively.

Table 2. Result for the tests of the cohorts for each group

Self-recognition test													
Score Exp.	Group A			Group B			Group C			Group D			Correct (%)
	CN	MCI	AD										
CN	128	8	-	197	12	3	142	11	6	40	-	-	507 (92)
MCI	16	139	-	74	155	33	59	120	30	5	88	3	502 (69)
AD	-	-	13	6	1	60	7	2	75	1	3	79	227 (91)

Stratified 3-fold cross validation test													
Score Exp.	Group A			Group B			Group C			Group D			Correct (%)
	CN	MCI	AD										
CN	122	14	-	177	28	7	133	20	6	31	6	3	463 (84)
MCI	33	122	-	97	130	35	68	103	38	9	75	12	430 (59)
AD	1	-	12	5	9	53	7	11	66	2	8	73	204 (82)

"Exp." column, outside of the parenthesis, represents the number of MR images base on the clinical test, and "Score" row, inside of the parenthesis, represents the number of MR images base on our test using score matrix. AD, Alzheimer's disease; CN, cognitively normal; MCI, mild cognitive impairment.

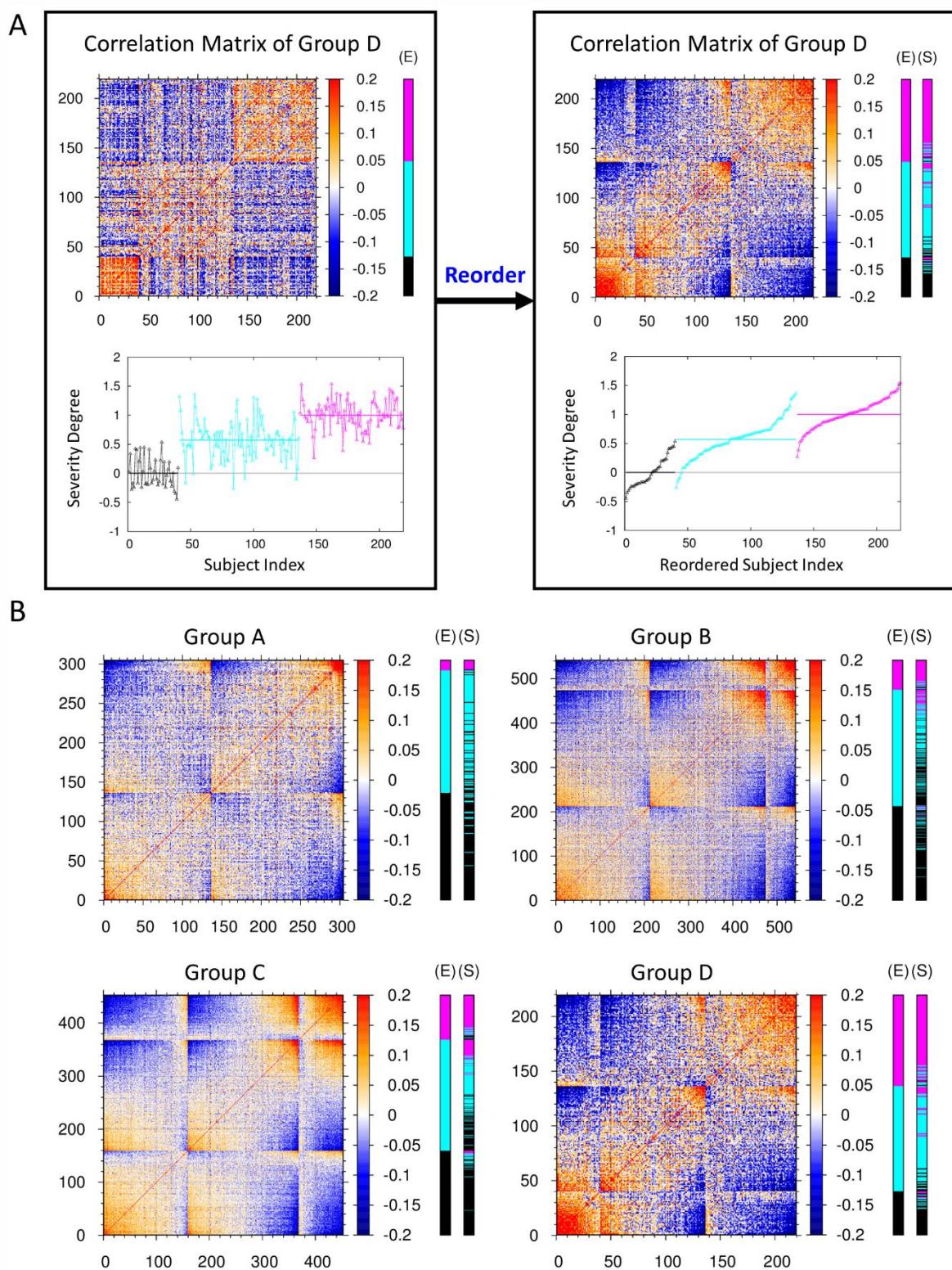
1

2 Estimating the severity of AD of subjects by covariance correlation matrix

3 Developing a quantitative measure to tell how serious the degree of dementia
 4 progression for a given subject is very important for diagnosing and clinically treating
 5 patients with MCI and AD with the different degree of Alzheimer dementia. In this study,
 6 we already identified essential ROI vertices and constructed the statistical score
 7 matrices as a classifier ensuring the correct prediction of subjects with AD at more

1 than 80% accuracy that they belong to the AD cohort. Thus, we extracted the cortical
2 thickness profile (or vector) at essential ROI vertices for all brain images, and
3 constructed the covariance correlation matrix between them. Then we compared the
4 profile for a given subject's image with that of patients with AD, to estimate the severity
5 degree of AD for a given subject towards patients with AD (for the detailed method,
6 see the section 2.4 and Figure 3A). The personalized quantitative severity degree of
7 dementia (see the equation (8)) is plotted at the right-bottom graph of Fig. 3A for each
8 subject of CN, MCI, AD cohort of the group D in the ascending order. The values of
9 severity degree of dementia were distributed around the averaged value of 0 (ranging
10 from about -0.5 to +0.5) for subjects with CN, 0.5 (ranging from about -0.2 to +1.2) for
11 subjects with MCI, and 1.0 (ranging from about +0.2 to +1.5) for subjects with AD,
12 respectively. The distribution of the severity degree for subjects with MCI was laid
13 across both ranges of those for CN and AD, which points out that this is the intrinsic
14 source of the low success ratio in determining the dementia cohort of subjects with
15 MCI. One could sort out quantitatively the broad spectrum of the severity of dementia
16 for MCI subjects in that whether they are prone to CN or how much they are
17 progressed toward AD. Given a new person for diagnosing the dementia state, one of
18 the cohort CN, MCI, AD was assigned by the equation (3) and the personalized
19 quantitative severity degree of dementia was estimated by the equation (8). Then, with
20 these two-qualitative and quantitative-determinants, one may infer that a new person
21 with the estimated severity degree below 0.0 is most likely to be CN, with that between
22 0.0 and 0.5 might be CN or MCI prone to CN, with that between 0.5 and 1.0 might be
23 MCI prone to AD or AD, and with that above 1.0 is most likely to be AD state.
24 We constructed the covariance correlation matrices for all groups A, B, C, D of the
25 average cortical thickness and observed the common pattern in the matrices that

1 subjects with AD (CN), possessing a strong correlation among themselves are
2 clustered at the top-right (bottom-left) corner, represented by the cluster of red colors
3 (Figure S4). Also, we presented the reordered covariance correlation matrices by the
4 severity degree of AD and the results to which one of the CN, MCI, AD cohorts each
5 human brain would belong, based on both the clinical test and our stratified 3-fold
6 cross validation test (Figure 3B). After comparing the result from our independent
7 validation test with that of the clinical test, we noted that those subjects which were
8 predicted to belong to the MCI cohort by the clinical test and yet estimated to have the
9 higher (lower) severity degree of AD by our estimation, were predicted to belong to the
10 AD (CN) cohort from the our validation test.



1 **Figure 3. Covariance correlation matrix and severity degree of AD.** (A) The left-top
2 heat map is the covariance correlation matrix for group D of average cortical thickness. The
3 x- and y-axes represent the indices 1 to 40 for CN subjects, 41 to 136 for MCI subjects,
4

1 and 137 to 219 for AD subjects. Here, red (blue) color represents the high (low) correlation
2 between two subjects at the essential ROI vertices. The extra E-cohort color bar at the right
3 of the heat map represents the clinically determined cohort of CN subjects and subjects
4 with MCI and AD denoted by black, cyan, and magenta colors, respectively. The left-bottom
5 graph illustrates the personalized severity degree of dementia for each subjects of the
6 group D in terms of a quantitative value, ranging from 0 for the basin of CN state to 1 for
7 the basin of AD state (for the detailed method, see the section 2.4). The average values of
8 this severity degree in each cohort are denoted by horizontal lines, respectively. The left
9 panel is reordered into the right panel according to the ascending value of the severity
10 degree in each cohort. For those subjects with MCI, the distribution of severity degree of
11 dementia is very broad. One can sort out the broad spectrum of the severity of dementia
12 for MCI subjects in that whether they are prone to CN or how much they are progressed
13 toward AD. (B) The reordered covariance correlation matrices for A, B, C, and D groups of
14 average cortical thickness together with the determination of CN (black), MCI (cyan), and
15 AD (magenta) cohorts by clinical (E-cohort color bar) exam and by the stratified 3-fold cross
16 validation test of this study (S-cohort color bar). The original covariance correlation matrices
17 for the four groups of average cortical thickness are provided in Figure S4.

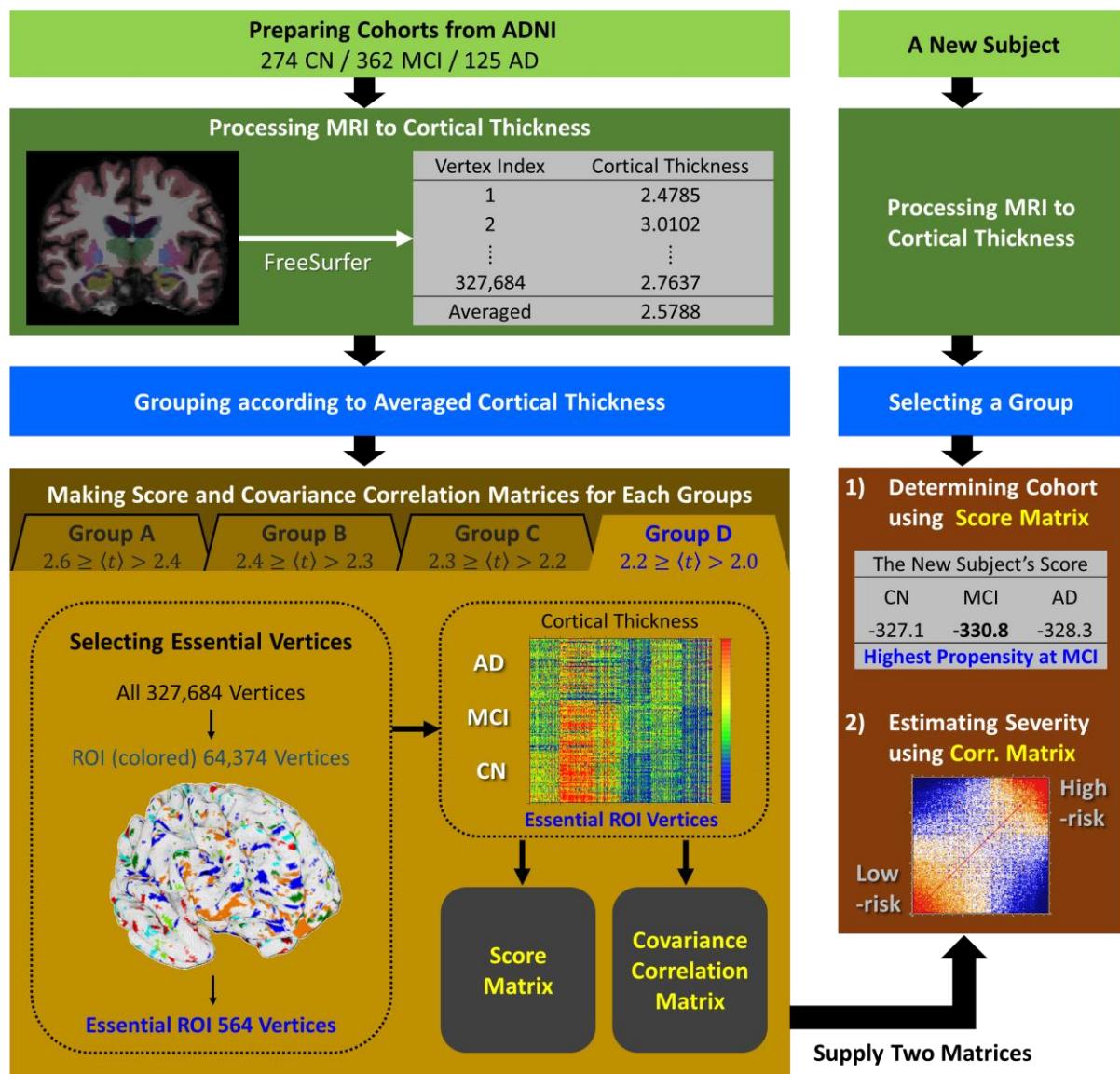
18

19

1 Discussion

2 Based on the cortical thickness data of 1516 human brain images from the ADNI data
3 set, we presented a new algebraic approach for both (1) the identification of the cohort
4 (CN, MCI, AD) a given subject would belong to and (2) the quantitative estimation of
5 the severity degree of AD for a given new person (Figure 4). A total of 1516 human
6 brain MR images were partitioned into four groups by the average cortical thickness
7 of each subject. Out of 327,684 vertices on the cortex, a few hundred essential ROI
8 vertices for each group were identified, which were enough to distinguish the cortical
9 thickness distribution of the CN, MCI, and AD cohorts from each other. Statistical score
10 matrices using the cortical thickness on the essential ROI vertices were constructed
11 as a classifier for determining the cohort of a given subject. Out of 547 CN, 722 MCI,
12 and 247 AD subjects predetermined clinically, the success ratio for recognizing their
13 cohort was 92% with CN, 69% with MCI, and 91% with AD subjects. On the other hand,
14 the stratified 3-fold cross-validation test gave the correct prediction rate of 84% with
15 CN, 59% with MCI, and 82% in subjects with AD; this is in agreement with the results
16 of clinical determination. Using the quantitative severity degree of AD for subjects, we
17 could explain the reason why the inevitable uncertainty in the determination of the MCI
18 cohort arose by the very broad distribution of the severity degree of dementia which
19 MCI subjects possess intrinsically. We suggested that the severity degree of AD
20 presented in this study would be a realistic measure for the quantitative personalized
21 diagnosis of a given subject instead of tri-partitioning the classification of a subject's
22 cohort only by CN, MCI or AD. It is the continuous degree for a given subject along
23 the scale from 0 for the basin of CN state to 1 for the basin of AD state. One could sort
24 out quantitatively the broad spectrum of the severity degree of dementia for MCI or AD
25 subjects with the different severity degree of dementia in that whether they are prone

1 to CN or how much they are progressed toward AD. This study not only provided a
2 straightforward algebraic approach to analyzing the cortical thicknesses of human
3 brains but also suggested quantitative measures by which one could estimate both the
4 cohort and the severity degree of AD for a given new subject based on the neuro-
5 images from the structural MRI. The MRI data of a larger number of human brains
6 could also be implemented into this study in a systematic and robust manner, which
7 would facilitate the better diagnose of AD with the different degree of dementia.



8
9 **Figure 4. The flow chart for the determination of cohorts and the estimation of the**

1 **severity degree of AD.** The left shows the process of constructing the score matrix and
2 covariance correlation matrix from the cortical thickness big data of subjects derived from
3 the ADNI data set. The right shows the process of determining the cohort and the severity
4 degree of AD for a new given subject.

5

1 Acknowledgments

2 This work was supported by the National Research Foundation, Korea [grant number
3 NRF 2017R1E1A1A03070854]. We acknowledged DGIST supercomputing big data
4 center for the allocation of supercomputing resources. We appreciated Mookyung
5 Cheon and Wookyung Yu for the fruitful discussions, and also Keonho Lee and
6 Jangjae Lee of Chosun university for the preparation of MRI image data in the initial
7 stage of this work.

8 Data collection and sharing for this project was funded by the AD Neuroimaging
9 Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI
10 (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the
11 National Institute on Aging, the National Institute of Biomedical Imaging and
12 Bioengineering, and through generous contributions from the following: AbbVie,
13 Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech;
14 BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate;
15 Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; Euroimmun; F.
16 Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE
17 Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development,
18 LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity;
19 Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research;
20 Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal
21 Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The
22 Canadian Institutes of Health Research is providing funds to support ADNI clinical
23 sites in Canada. Private sector contributions are facilitated by the Foundation for the
24 National Institutes of Health (www.fnih.org). The grantee organization is the Northern
25 California Institute for Research and Education, and the study is coordinated by the

1 Alzheimer's Therapeutic Research Institute at the University of Southern California.
2 ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of
3 Southern California.

4

5 **Competing interests**

6 All authors report no competing interests.

7

1 References

- 2 DALE, A. M., FISCHL, B. & SERENO, M. I. 1999. Cortical surface-based analysis. I. Segmentation and
3 surface reconstruction. *Neuroimage*, 9, 179-94.
- 4 FISCHL, B., SERENO, M. I. & DALE, A. M. 1999. Cortical surface-based analysis. II: Inflation, flattening,
5 and a surface-based coordinate system. *Neuroimage*, 9, 195-207.
- 6 GALVIN, J. E., SADOWSKY, C. H. & NINCDS, A. 2012. Practical guidelines for the recognition and
7 diagnosis of dementia. *J Am Board Fam Med*, 25, 367-82.
- 8 HARTIKAINEN, P., RASANEN, J., JULKUNEN, V., NISKANEN, E., HALLIKAINEN, M., KIVIPELTO, M.,
9 VANNINEN, R., REMES, A. M. & SOININEN, H. 2012. Cortical thickness in frontotemporal
10 dementia, mild cognitive impairment, and Alzheimer's disease. *J Alzheimers Dis*, 30, 857-74.
- 11 HOJJATI, S. H., EBRAHIMZADEH, A. & BABAJANI-FEREMI, A. 2019. Identification of the Early Stage
12 of Alzheimer's Disease Using Structural MRI and Resting-State fMRI. *Front Neurol*, 10, 904.
- 13 IM, K., LEE, J. M., SEO, S. W., YOON, U., KIM, S. T., KIM, Y. H., KIM, S. I. & NA, D. L. 2008. Variations
14 in cortical thickness with dementia severity in Alzheimer's disease. *Neurosci Lett*, 436, 227-
15 31.
- 16 KIM, B. H., CHOI, Y. H., YANG, J. J., KIM, S., NHO, K., LEE, J. M. & ALZHEIMER'S DISEASE
17 NEUROIMAGING, I. 2020. Identification of Novel Genes Associated with Cortical Thickness
18 in Alzheimer's Disease: Systems Biology Approach to Neuroimaging Endophenotype. *J
19 Alzheimers Dis*.
- 20 KIM, H. J., YE, B. S., YOON, C. W., NOH, Y., KIM, G. H., CHO, H., JEON, S., LEE, J. M., KIM, J. H., SEONG,
21 J. K., KIM, C. H., CHOE, Y. S., LEE, K. H., KIM, S. T., KIM, J. S., PARK, S. E., KIM, J. H., CHIN, J.,
22 CHO, J., KIM, C., LEE, J. H., WEINER, M. W., NA, D. L. & SEO, S. W. 2014. Cortical thickness
23 and hippocampal shape in pure vascular mild cognitive impairment and dementia of
24 subcortical type. *Eur J Neurol*, 21, 744-51.
- 25 LEBEDEV, A. V., WESTMAN, E., BEYER, M. K., KRAMBERGER, M. G., AGUILAR, C., PIRTOSEK, Z. &
26 AARSLAND, D. 2013. Multivariate classification of patients with Alzheimer's and dementia
27 with Lewy bodies using high-dimensional cortical thickness measurements: an MRI surface-
28 based morphometric study. *J Neurol*, 260, 1104-15.
- 29 PATERNICO, D., MANES, M., PREMI, E., COSSEDDU, M., GAZZINA, S., ALBERICI, A., ARCHETTI, S.,
30 BONOMI, E., COTELLI, M. S., COTELLI, M., TURLA, M., MICHELI, A., GASPAROTTI, R., PADOVANI,
31 A. & BORRONI, B. 2016. Frontotemporal dementia and language networks: cortical thickness
32 reduction is driven by dyslexia susceptibility genes. *Sci Rep*, 6, 30848.
- 33 QIU, S., JOSHI, P. S., MILLER, M. I., XUE, C., ZHOU, X., KARJADI, C., CHANG, G. H., JOSHI, A. S., DWYER,
34 B., ZHU, S., KAKU, M., ZHOU, Y., ALDERAZI, Y. J., SWAMINATHAN, A., KEDAR, S., SAINT-
35 HILAIRE, M. H., AUERBACH, S. H., YUAN, J., SARTOR, E. A., AU, R. & KOLACHALAMA, V. B.
36 2020. Development and validation of an interpretable deep learning framework for
37 Alzheimer's disease classification. *Brain*.
- 38 QUERBES, O., AUBRY, F., PARIENTE, J., LOTTERIE, J. A., DEMONET, J. F., DURET, V., PUEL, M., BERRY, I.,

- 1 FORT, J. C., CELSIS, P. & ALZHEIMER'S DISEASE NEUROIMAGING, I. 2009. Early diagnosis of
2 Alzheimer's disease using cortical thickness: impact of cognitive reserve. *Brain*, 132, 2036-
3 47.
- 4 RAJ, A., KUCEYESKI, A. & WEINER, M. 2012. A network diffusion model of disease progression in
5 dementia. *Neuron*, 73, 1204-15.
- 6 SOLOMON, A. & SOININEN, H. 2015. Dementia: Risk prediction models in dementia prevention. *Nat
7 Rev Neurol*, 11, 375-7.
- 8 TETREAUULT, A. M., PHAN, T., ORLANDO, D., LYU, I., KANG, H., LANDMAN, B., DARBY, R. R. &
9 ALZHEIMER'S DISEASE NEUROIMAGING, I. 2020. Network localization of clinical, cognitive,
10 and neuropsychiatric symptoms in Alzheimer's disease. *Brain*, 143, 1249-1260.
- 11 YU, W., LEE, W., LEE, W., KIM, S. & CHANG, I. 2011. Uncovering symmetry-breaking vector and
12 reliability order for assigning secondary structures of proteins from atomic NMR chemical
13 shifts in amino acids. *J Biomol NMR*, 51, 411-24.
- 14
- 15

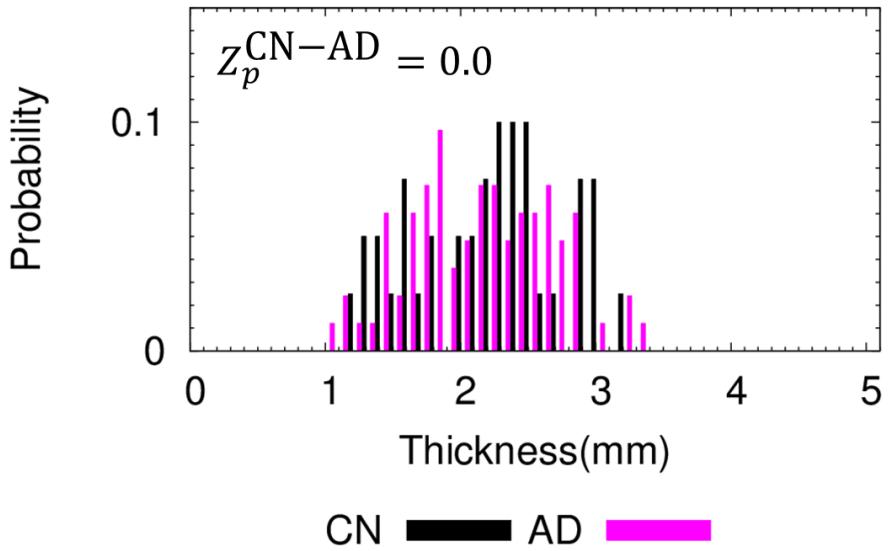
1 Supplementary materials

Supplementary Table 1. Upper / lower cut-off Z score and number of essential ROI vertices

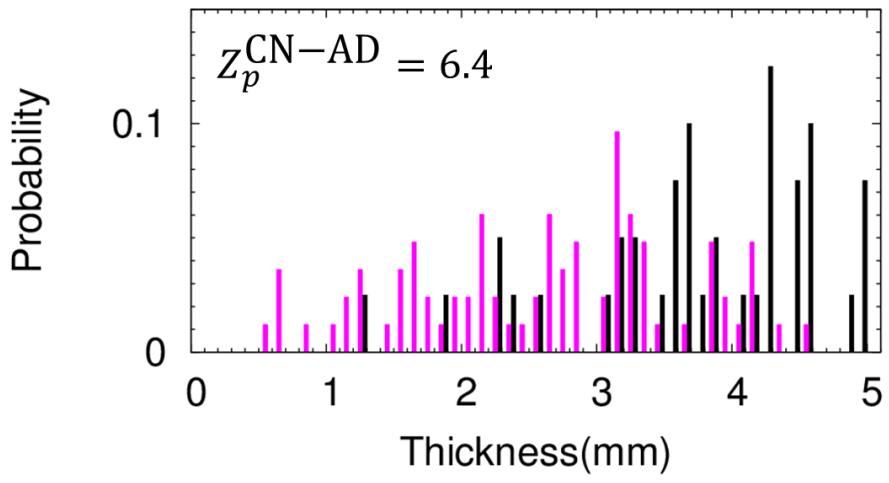
	$Z^{\text{CN-MCI}}$	$Z^{\text{CN-AD}}$	$Z^{\text{MCI-AD}}$	Number of essential ROI vertices
Group A	4.4 / -4.1	6.3 / -5.0	6.4 / -5.1	479
Group B	4.7 / -3.9	6.1 / -4.4	4.8 / -4.3	520
Group C	4.9 / -4.4	8.7 / -4.3	5.7 / -4.2	494
Group D	4.2 / -4.2	5.5 / -4.6	4.7 / -4.2	564

AD, Alzheimer's disease; CN, cognitively normal; MCI, mild cognitive impairment; ROI, region-of-interest.

A



B



1

CN ■■■ AD ■■■■■

2

Supplementary Figure 1. The distribution of cortical thickness of CN and AD subjects

3

in group D of average cortical thickness. (A) At the smallest $|Z_p^{CN-AD}|$ value, the two

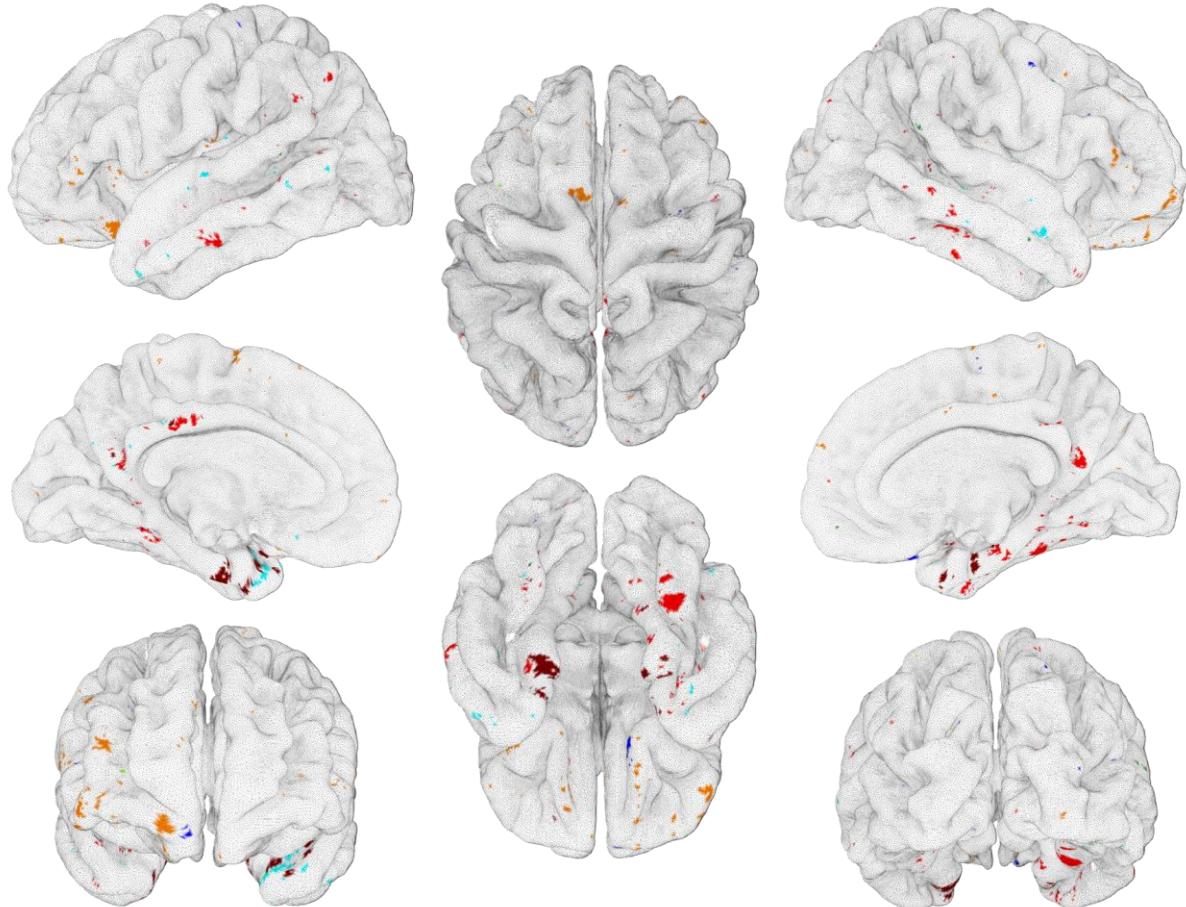
4

distribution curves of CN and AD subjects are not distinguished. On the other hand, (B) at

5

the largest $|Z_p^{CN-AD}|$ value, the two distribution curves are distinguished.

6

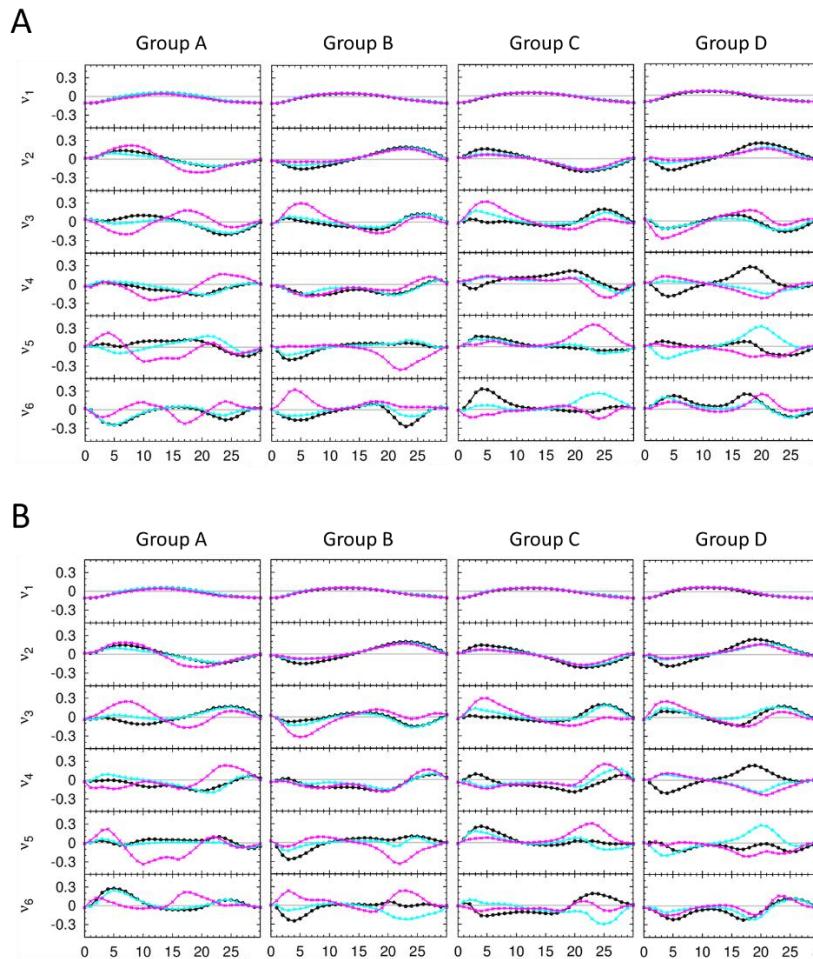


1
2 **Supplementary Figure 2. ROI vertices found commonly more than three times from**
3 **A, B, C, and D groups of the average cortical thickness.** ROI vertices at which the
4 cortical thickness of CN subjects are thicker (thinner) are represented by cyan (blue) color.
5 ROI vertices at which the cortical thickness of subjects with MCI are thicker (thinner) are
6 represented by green (dark green) color. ROI vertices at which the cortical thickness of
7 subjects with AD are thicker (thinner) are represented by orange (red) color. The ROI
8 vertices at which the cortical thickness decreases in the descending order of CN-MCI-AD
9 are represented by dark red.

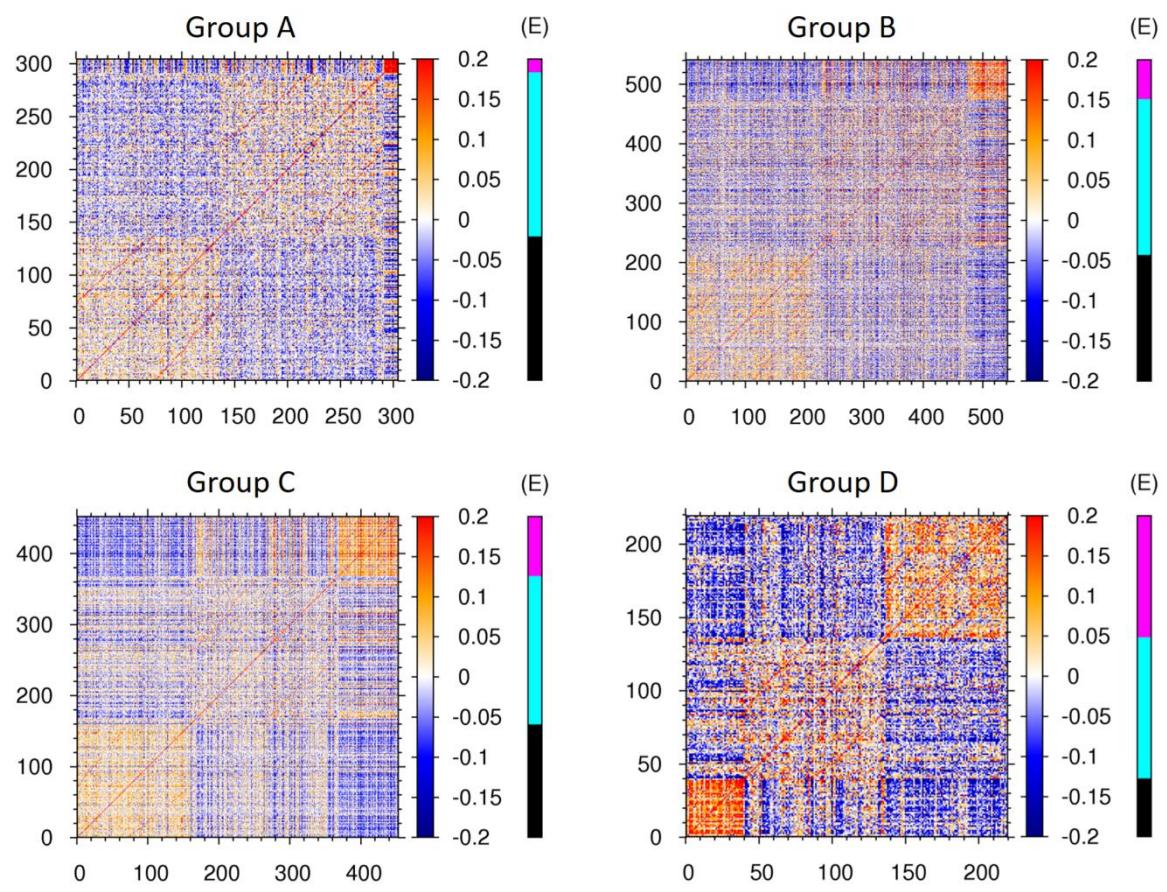
10

11

12



Supplementary Figure 3. The character of score matrix for each group of average cortical thickness. (A) The results of singular value decomposition analysis on score matrices, which are composed of 365 CN, 481 MCI, 165 AD human brain images as a training set and used for the second iteration of stratified 3-fold cross validation test. And (C) the results of that which are composed of 366 CN, 483 MCI, 166 AD human brain images as a training set and used for the third iteration of stratified 3-fold cross validation test. For each group of average cortical thickness, six singular vectors corresponding to the six largest singular values are presented. Here, x-axis is m value defined in the section 2.3, and y-axis is an arbitrary unit for the singular vectors. For each graph, the singular vector components for CN, MCI, and AD subjects are plotted by black, cyan, magenta colors, respectively.



Supplementary Figure 4. Covariance correlation matrices for each group of cortical thickness. The heat maps is the covariance correlation matrix for group A to D of average cortical thickness. The extra E-cohort color bar at the right of the heat maps represent the clinically determined cohort of cognitively normal subjects and subjects with mild cognitive impairment and Alzheimer's disease denoted by black, cyan, and magenta colors, respectively.