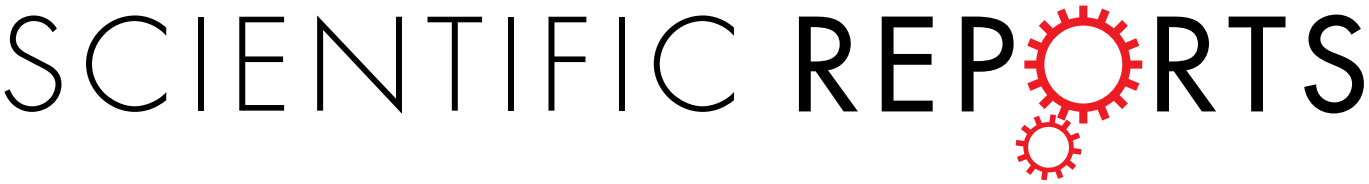

OPEN

Mining Outcome-relevant Brain Imaging Genetic Associations via Three-way Sparse Canonical Correlation Analysis in Alzheimer's Disease

Received: 16 August 2016
Accepted: 07 February 2017
Published: 14 March 2017

Xiaoke Hao[1], Chanxiu Li[1], Lei Du[2], Xiaohui Yao[3], Jingwen Yan[3,4], Shannon L. Risacher[3], Andrew J. Saykin[3], Li Shen[3], Daoqiang Zhang[1] & Alzheimer's Disease Neuroimaging Initiative[†]

Neuroimaging genetics is an emerging field that aims to identify the associations between genetic variants (e.g., single nucleotide polymorphisms (SNPs)) and quantitative traits (QTs) such as brain imaging phenotypes. In recent studies, in order to detect complex multi-SNP-multi-QT associations, bi-multivariate techniques such as various structured sparse canonical correlation analysis (SCCA) algorithms have been proposed and used in imaging genetics studies. However, associations between genetic markers and imaging QTs identified by existing bi-multivariate methods may not be all disease specific. To bridge this gap, we propose an analytical framework, based on three-way sparse canonical correlation analysis (T-SCCA), to explore the intrinsic associations among genetic markers, imaging QTs, and clinical scores of interest. We perform an empirical study using the Alzheimer's Disease Neuroimaging Initiative (ADNI) cohort to discover the relationships among SNPs from AD risk gene *APOE*, imaging QTs extracted from structural magnetic resonance imaging scans, and cognitive and diagnostic outcomes. The proposed T-SCCA model not only outperforms the traditional SCCA method in terms of identifying strong associations, but also discovers robust outcome-relevant imaging genetic patterns, demonstrating its promise for improving disease-related mechanistic understanding.

Alzheimer's disease (AD) is the most common form of dementia characterized by progressive impairment of memory and other cognitive functions in people over 65 years of age[1]. It is an important research topic to develop methods for early diagnosis of AD. At present, many studies have focused on searching for biomarkers from brain imaging data as well as molecular and cellular data to investigate the pathological changes[2]. To further aid the development of effective diagnostic and therapeutic approaches, it has received increasing attention to study AD at the system biology level. For example, revealing the biological pathway from the microcosmic genetic factors to macroscopic brain anatomy has the potential to understand pathogenicity mechanisms underlying the disordered cognition and behavior.

Enabled by recent advances in high-throughput genotyping and multimodal neuroimaging technologies, imaging genetics is becoming an emerging research field for discovering the associations between genetic markers such as single nucleotide polymorphisms (SNPs) and quantitative traits (QTs) extracted from structural or functional neuroimaging data[3,4]. Thus, it holds great promise for us to understand the complex neurogenetic and neurobiological mechanism of complex brain disorders[5].

In prior imaging genetics studies, univariate and multivariate regression methods have been typically used to capture the effective associations between SNPs and neuroimaging data[6–10]. More recently, bi-multivariate

[1]College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China. [2]School of Automation, Northwestern Polytechnical University, Xi'an 710072, China. [3]Department of Radiology and Imaging Sciences, School of Medicine, Indiana University, Indianapolis, IN 46202, USA. [4]School of Informatics and Computing, Indiana University, Indianapolis, IN 46202, USA. [†]A comprehensive list of consortium members appears at the end of the paper. Correspondence and requests for materials should be addressed to L.S. (email: shenli@iu.edu) or D.Z. (email: dqzhang@nuaa.edu.cn)

analysis techniques such as various structured sparse canonical correlation analysis (SCCA) models have attracted increasing attention in brain imaging genetics to detect complex multi-SNP-multi-QT associations[11–18]. The essence of all the structured SCCA approaches, incorporating valuable prior knowledge, is to find the best linear transformations for imaging and genetic features respectively so that the strongest correlation between the imaging and genetic components can be achieved. These methods have the potential to discover effective imaging genetic associations, while the identified genotypic and phenotypic markers may not be disease specific. To overcome this limitation, in this paper, we focus on exploring three-way associations among genetic markers, imaging QTs, and cognitive and diagnostic outcomes. Our goal is to reveal the relationships among these multidimensional genetics, imaging and outcome data, and contribute to a better understanding of pathogenicity mechanisms in AD.

Of note, some diagnosis information guided methods have been proposed in the field of imaging genetics. The sRRR model proposed in[19,20] used a two-step procedure for detecting genetic factors associated with disease relevant imaging phenotypes by using penalized linear discriminant analysis. More recently, a Bayesian framework was used to select the relevant features along the pathway from gene to imaging and then to symptom[21]. It is worth noting that both models treated diagnosis information as binary status (e.g., AD and normal control (NC)) for imaging genetic association studies. Actually, in the spectrum between NC and AD, there exist other progressive stages. For example, in the Alzheimer's Disease Neuroimaging Initiative (ADNI) study, there were participant groups labeled as Significant Memory Concern (SMC), Early Mild Cognitive Impairment (EMCI) and Late Mild Cognitive Impairment (LMCI). In addition, the cognitive scores such as Mini-Mental State Examination (MMSE), Clinical Dementia Rating (CDR), ADNI Memory (ADNI-MEM) and ADNI Executive Functioning (ADNI-EF), which are neuropsychological assessment measures from different aspects, are often used as quantitative descriptions of symptom severity instead of binary diagnosis. Accordingly, identification of imaging genetic associations relevant to these diagnostic and cognitive outcomes may yield important information for a better understanding of disease-specific mechanisms.

With these observations, we consider the outcome-relevant imaging genetic association study as a multi-view multivariate correlation problem, which can be solved by CCA and partial least squares (PLS), as well as their sparse versions (including SCCA and SPLS)[11–13,22,23]. Thus, following the existing imaging genetic studies via bi-multivariate SCCA[11], we propose a three-way sparse canonical correlation analysis (T-SCCA) framework to explore the intrinsic associations among SNP loci, neuroimaging features and clinical score outcomes. Specifically, in this study, the outcomes of interest include cognitive scores (CS) and diagnosis status (DS). We evaluate the effectiveness of the proposed method by identifying three-way associations among 85 candidate SNPs from the top AD risk gene *APOE*, 116 imaging QTs extracted from structural magnetic resonance imaging (MRI) scans, and relevant cognitive and diagnostic outcomes, using the Alzheimer's Disease Neuroimaging Initiative (ADNI) data as a test bed. The experimental results demonstrate that the proposed T-SCCA model not only outperforms the standard two-way SCCA method in terms of identifying strong associations, but also discovers robust outcome-relevant imaging genetic patterns, demonstrating its promise for improving disease-related mechanistic understanding.

Imaging genetic associations

Imaging genetic associations via bi-multivariate analysis. We first describe relevant notation. We use lowercase letters to denote vectors, and uppercase letters to denote matrices. For a given matrix $M = \left[m_j^i\right]$, we denote its i-th row and j-th column as m^i and m_j respectively. Let $X=[x_1,\ldots,x_n]^T \in R^{n\times p}$ be the SNP genotype data, $Y=[y_1, \ldots, y_n]^T \in R^{n\times q}$ be the imaging QT data (i.e., voxel-based morphometry measures in this work), where n is the number of participants, and p and q are the number of SNPs and QTs, respectively.

For detecting complex multi-SNP-multi-QT associations, sparse canonical correlation analysis (SCCA)[11–13] seeks linear transformations of variables X and Y to achieve the maximal correlation between Xw_1 and Yw_2 by introducing penalty terms simultaneously, which can be formulated as:

$$\max_{w_1,w_2} w_1^T X^T Y w_2$$
$$s.\ t.\ \ w_1^T X^T X w_1 \le 1,\ \ w_2^T Y^T Y w_2 \le 1,\ \ ||w_1||_1 \le c_1,\ \ ||w_2||_1 \le c_2 \quad (1)$$

where w_1 and w_2 are canonical loadings or weights, reflecting the contribution of each feature in the identified canonical correlation. Note that $w_1^T X^T X w_1 \le 1,\ \ w_2^T Y^T Y w_2 \le 1$ are used to embrace the covariance structure of the data in the model. $||w_1||_1 \le c_1$, $||w_2||_1 \le c_2$ are constraints for controlling the sparsity so that only a small number of relevant features will be selected automatically from the SNP and imaging data.

Cognitive score (CS)-relevant imaging genetic associations via three-way SCCA. In this study, for revealing the biological mechanism specific to the disease, we aim to discover imaging genetic associations that are relevant to cognitive scores (CSs) or diagnosis status. The first attempt focuses on involving multi-assessment CSs (i.e., MMSE, CDR, ADNI-MEM and ADNI-EF). Let $Z=[z_1, \ldots, z_n]^T \in R^{n\times r}$ be the CSs, where n is the number of participants, r is the number of CSs. Since different CSs can provide complementary perspectives on neuropsychological assessments, we aim to seek a set of linear transforms to estimate the contribution of each individual cognitive score in imaging genetic associations.

The formulation of CS-relevant imaging genetic associations can be extended from eq. (1) as follows:

Method	Correlation Coefficient on Training Set					
	F1	F2	F3	F4	F5	Mean + Std
BM-SCCA	0.2619	0.2810	0.1846	0.2679	0.2755	0.2542 ± 0.0396
CS-SCCA	0.3436	0.3819	0.3743	0.3536	0.3798	0.3667 ± 0.0171
DS-SCCA	0.3519	0.3843	0.3848	0.3584	0.3822	0.3723 ± 0.0159

Table 1. 5-fold cross-validation results on ADNI: The model learned from the training data is used to estimate the correlation coefficients on the training set.

Method	Correlation Coefficient on Test Set					
	F1	F2	F3	F4	F5	Mean + Std
BM-SCCA	0.1996	0.1848	−0.0250	0.2845	0.2320	0.1752 ± 0.1183
CS-SCCA	0.3328	0.2126	0.2258	0.3275	0.2180	0.2633 ± 0.0612
DS-SCCA	0.3566	0.2173	0.2200	0.3139	0.2474	0.2711 ± 0.0616

Table 2. 5-fold cross-validation results on ADNI: The model learned from the training data is used to estimate the correlation coefficients on the testing set.

$$\begin{aligned} &\max_{w_1,w_2,w_3} w_1^T X^T Y w_2 + w_1^T X^T Z w_3 + w_2^T Y^T Z w_3 \\ &s.\ t.\ \ w_1^T X^T X w_1 \le 1,\ \ w_2^T Y^T Y w_2 \le 1,\ \ w_3^T Z^T Z w_3 \le 1, \\ &\quad ||w_1||_1 \le c_1,\ \ ||w_2||_1 \le c_2,\ \ ||w_3||_1 \le c_3 \end{aligned} \tag{2}$$

where w_3 is the canonical loading, reflecting the contribution of each neuropsychological assessment in the identified canonical correlation. Similar to the existing constraints in eq. (1), $w_3^T Z^T Z w_3 \le 1$ is a newly added constraint to embrace the covariance structure, and $||w_3||_1 \le c_3$ is another newly added constraint for selecting a small number of CS that are related to both imaging and genetics measures.

Diagnosis status (DS)-relevant imaging genetic associations via three-way SCCA. Besides the CSs, the diagnosis status (DS) considered as a qualitative measurement can also describe the progressive stages of AD. Accordingly, we propose another outcome-relevant imaging genetic association that involves DS. Let $Z \in R^{n\times 1}$ be the DS labels for all participants. The DS-relevant formulation, similar to eq. (2), can be used for mining DS-relevant imaging genetic associations.

It is worth noting that the equation of DS-relevant imaging genetic association is a special case of eq. (2) if $r = 1$. Similar to eq. (2), w_3 is the canonical loading reflecting the contribution of DS in the identified canonical correlation, so that the optimization of w_1,w_2 and w_3 can be also solved by the proposed algorithm.

Results

Characteristics of the dataset. The dataset comprises of 913 non-Hispanic Caucasian participants, including 211 normal control (NC), 82 significant memory concern (SMC), 273 early mild cognitive impairment (EMCI), 187 late mild cognitive impairment (LMCI) and 160 AD. In our experiments, we used the baseline structural MRI data with average voxel-based morphometry (VBM) measures in 116 regions of interest (ROIs), genotyping data with 85 candidate SNP loci, as well as cognitive scores (including MMSE, CDR, ADNI-MEM and ADNI-EF) and diagnosis status (including NC, SMC EMCI, LMCI and AD). For more details about the demographics and data processing, please see the complete information in the Methods section.

Experimental settings. In this imaging genetic association study, 5-fold cross-validation strategy is adopted to evaluate the effectiveness of our proposed method. For parameters of regularization, we determine their values by another nested 5-fold cross-validation on the training set. It is used to fine tune the parameters in the objective function in the range of $\{10^{-3}, 3\times10^{-3}, 10^{-2}, 3\times10^{-2}, 10^{-1}, 3\times10^{-1}, 1, 3, 10, 30, 100\}$. The parameters yielding the best performance in the inner cross-validation are finally used in the resulting model.

In the current experiments, we compare BM-SCCA (denoting conventional bi-multivariate SCCA), CS-SCCA (denoting cognitive scores-guided method via three-way SCCA), DS-SCCA (denoting diagnosis-guided method via three-way SCCA). Both CS-SCCA and DS-SCCA belong to the T-SCCA category.

Improved association between risk SNPs and phenotypic imaging markers. We compare our proposed T-SCCA methods (including CS-SCCA and DS-SCCA) with the conventional method (BM-SCCA). The performance on each dataset is assessed using the correlation coefficient (CC) between SNP and imaging data, which is widely used for association analysis measurements. The average results of CC across the 5-fold training and testing data are calculated respectively. As shown in Table 1 and Table 2, CS-SCCA and DS-SCCA yield the CC values of 0.2633 (0.3667) and 0.2711 (0.3723) on test (training) set, respectively, which are better than those of BM-SCCA. These results indicate that the search space of BM-SCCA could be too large such that the algorithm could converge to local optima without prior knowledge, while the regularizations of the restrictions

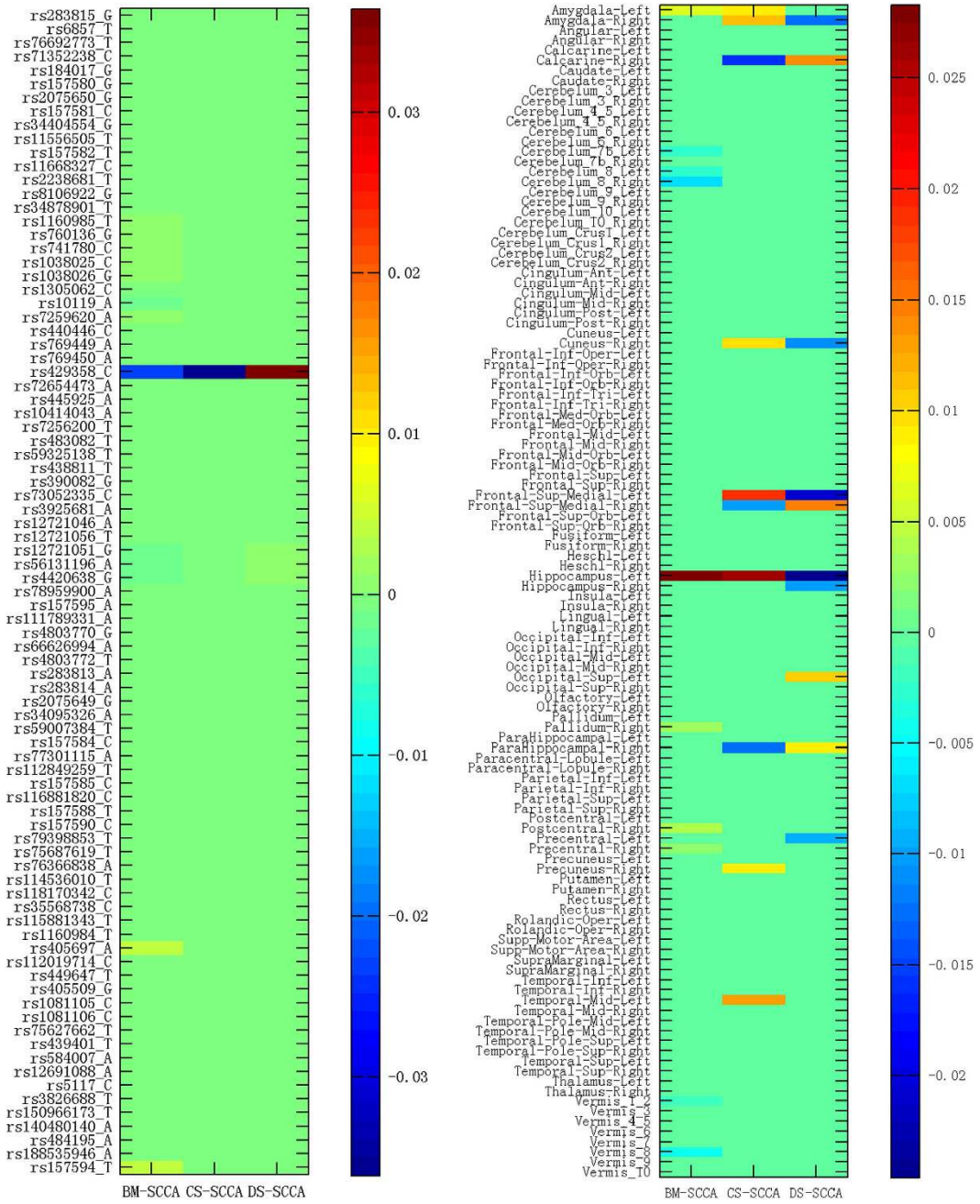

Figure 1. Heat map of average estimated canonical loadings on 85 *APOE* SNPs associated with 116 brain ROIs across 5-fold cross-validation respect to different methods.

on outcome-relevant information might be able to guide imaging genetic associations out of over-fitting. This demonstrates the disease information could help improve the performances of correlations between genotypes and imaging phenotypes.

In addition, we perform a permutation test using 1000 permutations with retraining each of BM-SCCA, SC-SCCA and DS-SCCA models to assess the statistical significance of the identified correlations on the test set. The p-value corresponds to the fraction of times that the correlation coefficient is greater or equal to the result from original data. The resulting p-values ($p < 0.001$) are statistically significant in all three cases, including BM-SCCA, CS-SCCA and DS-SCCA, respectively.

Besides improving correlation performances, one major goal of this study is to identify genotypic and phenotypic markers that are not only highly correlated to each other, but also relevant to cognitive or diagnostic outcomes specific to AD. Figure 1 shows the heat map of average estimated canonical loadings on 85 *APOE* SNPs and 116 brain ROIs by BM-SCCA, CS-SCCA and DS-SCCA, respectively. The weighted colors of selected SNPs and brain regions presented by canonical loadings indicate the contributions of the corresponding genetic and phenotypic markers.

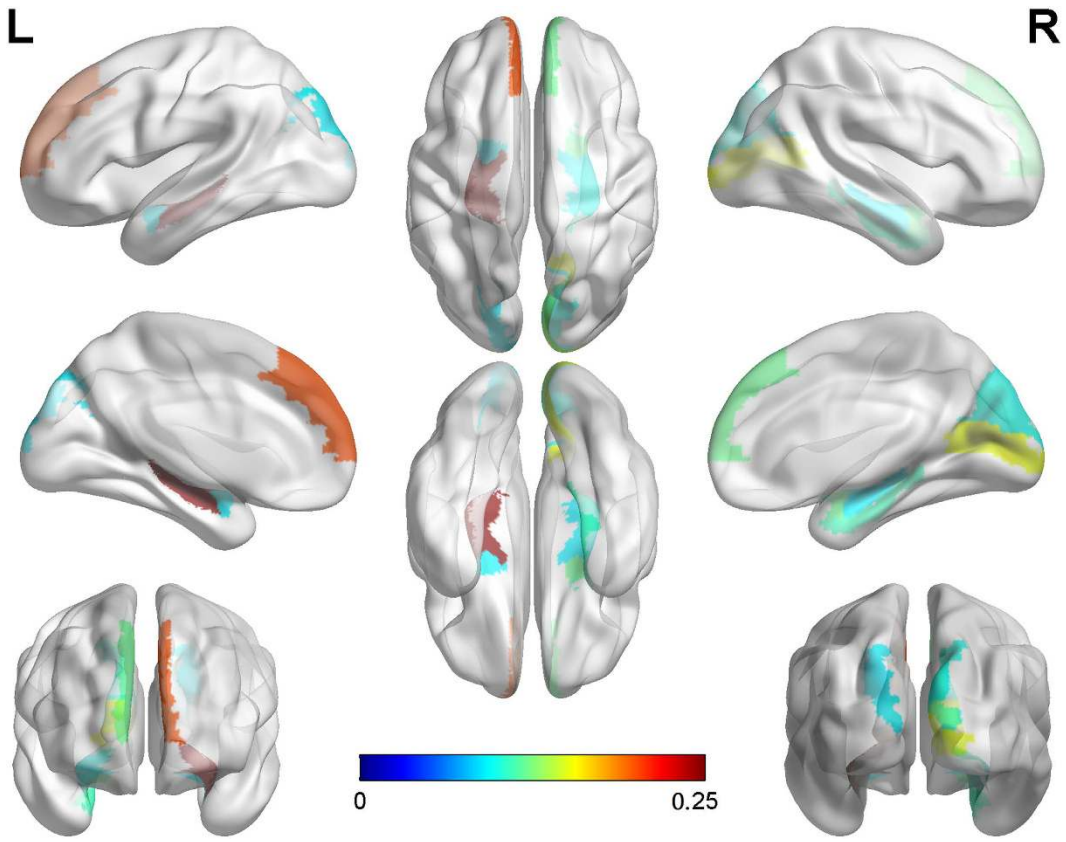

Figure 2. Visualization of mapping top 10 average estimated canonical loadings generated by T-SCCA (combination of CS-SCCA and DS-SCCA) onto the brain.

Discussion

As expected, the well-known locus rs429358 is identified to be associated with gray matter loss in multiple AD-relevant ROIs, which is in accordance with the previous studies[24]. The C allele increases the risk of AD in *APOE* e4, which is encoded by rs429358 (www.snpedia.com/index.php/APOE) (www.alzgene.org)[25]. BM-SCCA seems to yield other loci such as rs405697 (this genetic variant in APOE region has shown only to be associated with human longevity in the literature[26]) and rs157594.

For the phenotype identifications, besides the well-known AD-related ROI left hippocampus, signals in the cerebellum and vermis areas are detected by BM-SCCA. However, the morphometric changes of the cerebellum have not been widely validated as AD biomarkers. On the other hand, a few additional ROIs such as right amygdala, right calcarine, right cuneus, left and right frontal-sup-medial gyrus, right parahippocampal gurus have been detected as top 10 features associated with the risk genotype biomarker rs429358 by the proposed T-SCCA (including CS-SCCA and DS-SCCA). It's worth noting that CS-SCCA captures similar patterns on brain imaging as DS-SCCA does for the most part. More interestingly, the same ROIs selected simultaneously by two types of T-SCCA have similar weight values, which demonstrates the robust and consistent biomarker findings. Note that these weights have different signs, which are caused by the negative directionality of the cognitive score values in relation to the diagnosis values. The top 10 selected MRI-VBM imaging features, as well as their averaged estimated canonical loadings generated by T-SCCA (combination of SC-SCCA and DS-SCCA) across 5 cross-validation trials, are visualized in Fig. 2 by mapping them onto the human brain. The colors of the selected brain regions indicate the canonical loadings of the corresponding markers. The identified regions have potential clinical correlates in typical clinically well-described AD impairments. To our knowledge, the hippocampus is one of the first regions of the brain to suffer damage including memory loss and disorientation. In addition, amygdala atrophy is related to aberrant motor behavior, with potential relationships to anxiety and irritability[24]. Some existing results suggest that the magnitude of amygdala atrophy is comparable to that of the hippocampus in the earliest clinical stages of AD[27]. The analytical result reassures that our method identifies a well-known correlation between genotypes and phenotypes that is severely and consistently affected by pathology in AD. Besides confirming the prior findings, our method also yields the associations between *APOE* rs429538 and other eminent AD markers such as both left and right frontal-sup-medial gyrus. There also appear to be specific relationships among genotypes, phenotypes and neuropsychiatric symptoms that deserve further investigation.

As mentioned earlier, the quantitative CSs (MMSE, CDR, ADNI-MEM and ADNI-EF) used to index cognitive decline for disease severity are able to show the graded differences in participants, and we induce the sparsity constraint for selecting the related CSs in this imaging genetic associations study. The CS-SCCA yields the loading values of 0.0095, −0.0103, 0.0238, and −0.0026 on MMSE, CDR, ADNI-MEM and ADNI-EF, respectively.

It demonstrates ADNI-MEM is the top-ranked score that contributes to the multiple associations among gene, neuroimaging and cognition. Since all CSs are neuropsychological assessment measures from different aspects, the CS-SCCA provides a simple evaluating approach to investigate the relative contribution of each clinical score.

In summary, we have performed a neuroimaging genetics study for Alzheimer's disease (AD) to explore the relationship between genetic variations in the *APOE* gene and brain ROIs measured by voxel based morphometry (VBM). Various existing sparse canonical correlation analysis (SCCA) methods are only designed for bi-multivariate analysis, and often yield suboptimal results without considering the cognitive or diagnostic outcomes. With these observations, we have investigated a three-way sparse canonical correlation analysis (T-SCCA) framework to discover the multiple associations among SNP loci, neuroimaging features and phenotypic outcomes (including cognitive scores (CSs) and diagnosis status (DS)). The experimental results performed on 913 subjects from ADNI show that our proposed T-SCCA model can substantially achieve higher correlations between genotypic and phenotypic features. Specifically, besides the improved correlation performances, CS-SCCA captures similar patterns on canonical loadings as DS-SCCA. This supports the benefit of our general T-SCCA model that can also identify some significant and robust biomarkers in imaging genetic associations, revealing disease-specific patterns on the complex mechanisms. However, there could be different mechanisms leading to three-way associations between genetics, imaging and diagnosis. For example, on top of possible pathway from genetics to imaging and then to disease, part of the genetic influences on disease might not be mediated through features captured by neuroimaging. In addition, genetics might independently affect disease susceptibility and imaging features, resulting in association between imaging and diagnosis, or there might be hidden or confounding variables that drive these associations. Thus, it warrants further investigation to reveal the underlying mechanisms related to the three-way associations discovered by our proposed T-SCCA methods.

In this initial study, where the number of samples exceeds the number of total features, the T-SCCA model can be successfully applied for association discovery coupled with feature selection. However, if the dataset has more features than samples, this ill-conditioned problem can be addressed via dimensionality reduction or regularization. In particular, when the datasets contain far more features (e.g., SNPs at the genome-wide magnitude), it will greatly increase the computational complexity and memory requirement. Note that the normal equation in the optimization contains matrix inversion operations (the time complexity is $O(n^3)$, where n is the number of features). Therefore, it is an interesting future topic to develop a more efficient solution for our proposed T-SCCA and to identify potential markers from high-throughput genome-wide variants and neuroimaging quantitative traits in outcome-relevant imaging genetic studies.

In addition, in this study, we have explored the imaging genetic associations within a single population of non-Hispanic Caucasians. However, the effect of population structure is another important topic, and it may affect the identifications in multivariate associations due to the potential bias introduced by multiple populations in a study. In this case, the population structure adjustment should be considered in the study.

Methods

Subjects. Data used in the preparation of this study were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership by several organizations, including the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies, and non-profit organizations. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). The study protocols were approved by the institutional review boards of all participating centers (Nanjing University of Aeronautics and Astronautics, Northwestern Polytechnical University, Indiana University and ADNI (A complete list of ADNI sites is available at http://www.adni-info.org/.)) and written informed consent was obtained from all participants or authorized representatives. All the analytical methods were performed on the de-identified ADNI data, and were determined by Indiana University Human Subjects Office as IU IRB Review Not Required. In addition, these methods were carried out in accordance with the approved guidelines.

Participants were screened and enrolled according to criteria demonstrated in ADNI study protocol (http://adni-info.org/Scientists/ADNIStudyProcedures.html#). The general inclusion/exclusion criteria of the subjects from ADNI procedures manual (http://www.adni-info.org) are briefly described as follows:

1. NC participants have no subjective or informant-based complaint of memory decline and normal cognitive performance. The MMSE scores on NC should be between 24 and 30, CDR should be 0.
2. SMC participants have subjective memory concerns as assessed using the Cognitive Change Index (CCI; total score from first 12 items >16), no informant-based complaint of memory impairment or decline, and normal cognitive performance on the Wechsler Logical Memory Delayed Recall (LM-delayed) and the MMSE[28].
3. EMCI participants have a memory concern reported by the subject, informant, clinician, abnormal memory function approximately 1 standard deviation below normative performance adjusted for education level on the LM-delayed, an MMSE total score greater than 24.
4. Besides a subjective memory concern as reported by subject, study partner or clinician, CDR on LMCI subjects is 0.5 and Memory Box (MB) score must be at least 0.5.
5. MMSE score on AD should be between 20 and 26, and CDR should be 0.5 or 1.0.

In the practical diagnosis of AD, multiple clinical variables are generally acquired, e.g., MMSE, CDR, ADNI-MEM (composite score for memory) and ADNI-EF (composite score for executive functioning), etc.

Subjects	NC	SMC	EMCI	LMCI	AD
Number	211	82	273	187	160
Gender(M/F)	109/102	33/49	153/120	108/79	95/65
Age	76.14 ± 6.53	72.45 ± 5.67	71.48 ± 7.12	73.86 ± 8.44	75.18 ± 7.88
Education	16.45 ± 2.62	16.78 ± 2.67	16.08 ± 2.62	16.38 ± 2.81	15.86 ± 2.75
MMSE	29.01 ± 1.23	29.00 ± 1.22	28.38 ± 1.54	27.71 ± 1.73	24.00 ± 2.62
CDR	0.01 ± 0.07	0.00 ± 0.00	0.49 ± 0.08	0.49 ± 0.07	0.72 ± 0.27
ADNI-MEM	1.02 ± 0.58	1.12 ± 0.57	0.60 ± 0.60	0.07 ± 0.67	−0.76 ± 0.61
ADNI-EF	0.85 ± 0.69	0.73 ± 0.81	0.51 ± 0.74	0.18 ± 0.81	−0.53 ± 0.91

Table 3. Characteristics of the subjects. Note: NC = Normal Control, SMC = Significant Memory Concern, ECMI = Early Mild Cognitive Impairment, LCMI = Late Mild Cognitive Impairment, AD = Alzheimer's disease.

Specifically, MMSE is used to examine functions including registration, attention and calculation, recall, language, ability to follow simple commands and orientation[29,30]. CDR is a numeric scale used to assess a patient's cognitive and functional performance in six areas: memory, orientation, problem solving, community affairs, hobbies and personal care[31]. There are two derived composite scores for MEM and EF from ADNI. The formation of ADNI-MEM is complicated by the use of different word lists in the Rey Auditory Verbal Learning Test (RAVLT) and the ADAS-Cog, and by Logical Memory I data missing by design[32]. The formation of ADNI-EF includes Category Fluency-animals, Category Fluency-vegetables, Trails A and B, Digit span backwards, WAIS-R Digit Symbol Substitution, and 5 Clock Drawing items (circle, symbol, numbers, hands, time)[33]. The demographic information is summarized in Table 3.

SNP genotype data. Since genetic risk factors can help scientists focus on relevant biological pathways and form effective hypothesis for drug design, identifying risk genetic markers associated with brain imaging can help understand the underlying biological mechanisms. We downloaded the ADNI-GO/2 genotyping data, and performed quality control and population stratification using the approach described in the previous study[34]. To limit potential effects of population stratification, this study is focused only on analyzing non-Hispanic white participants. As the best-known genetic risk factor in AD, *APOE* (located on chromosome 19) has a key role in coordinating the mobilization and redistribution of cholesterol, phospholipids, and fatty acids, and it is implicated in mechanisms such as neuronal development, brain plasticity, and repair functions[35]. Thus, we focused our analysis on all SNPs within ±20 k base pairs of the *APOE* gene boundary based on the ANNOVAR (http://annovar.openbioinformatics.org) annotation, which include a total number of 85 SNPs as candidates. For the input in the models, each SNP value was coded in an additive fashion as 0, 1 or 2, indicating the number of minor alleles.

Imaging phenotype data. The MRI data used in this paper were also obtained from the ADNI database. We aligned the preprocessed imaging data (i.e., voxel based morphometry (VBM)) to each participant's same visit scan, and then created normalized gray matter density maps from the MRI data in the standard Montreal Neurological Institute (MNI) space as $2 \times 2 \times 2\,\mathrm{mm}^3$ voxels SPM software package[36]. 116 ROI level measurements of mean gray matter densities were further extracted based on the MarsBaR AAL atlas[37]. All measurements were pre-adjusted for age, gender, and education.

Objective function and algorithm design. In this section, we design an algorithm to solve the optimization problem defined in eq. (2). For the general formulation, using the Lagrange multiplier and writing the penalties into the matrix form, the objective function for mining CS-relevant or DS-relevant imaging genetic associations via three-way sparse canonical correlation analysis (T-SCCA) is as follows:

$$\begin{aligned} L(w_1, w_2, w_3) \;=\; & -w_1^T X^T Y w_2 - w_1^T X^T Z w_3 - w_2^T Y^T Z w_3 + \frac{1}{2}\,||Xw_1||_2^2 + \frac{1}{2}\,||Yw_2||_2^2 \\ & + \frac{1}{2}\,||Zw_3||_2^2 + \beta_1\,||w_1||_1 + \beta_2\,||w_2||_1 + \beta_3\,||w_3||_1 \end{aligned} \tag{3}$$

where $(\beta_1, \beta_2, \beta_3)$ and are the set of model parameters. Take the derivative regarding w_1, w_2 and w_3 separately and let them be zero:

$$w_1 = (X^T X + \beta_1 D_1)^{-1} X^T (Y w_2 + Z w_3) \tag{4}$$

$$w_2 = (Y^T Y + \beta_2 D_2)^{-1} Y^T (X w_1 + Z w_3) \tag{5}$$

$$w_3 = (Z^T Z + \beta_3 D_3)^{-1} Z^T (X w_1 + Y w_2) \tag{6}$$

where D_1 is a diagonal matrix with the k_1-th element as $\frac{1}{2\,||\,w_1^{k_1}\,||_1}$ $(k_1 \in [1, p])$, D_2 is a diagonal matrix with the k_2-th element as $\frac{1}{2\,||\,w_2^{k_2}\,||_1}$ $(k_2 \in [1, q])$, and D_3 is a diagonal matrix with the k_3-th element as $\frac{1}{2\,||\,w_3^{k_3}\,||_1}$ $(k_3 \in [1, r])$.

Since D_1 relies on w_1, D_2 relies on w_2, and D_3 relies on w_3, we introduce an iterative procedure to solve this objective. In each iteration, we first fix w_2 and w_3 to solve w_1, then fix w_1 and w_3 to solve w_2, and finally fix w_1 and w_2 to solve w_3. The procedure stops until it satisfies a predefined stopping criterion. Algorithm 1 shows the pseudo code of the T-SCCA algorithm for mining outcome-relevant imaging genetic associations.

Algorithm 1: T-SCCA for Mining Outcome-relevant Imaging Genetic Associations
Input: candidate SNPs $X=[x_1,\ldots,x_n]^T\in R^{n\times p}$,
neuroimaging ROIs $Y=[y_1,\ldots,y_n]^T\in R^{n\times q}$,
cognitive scores (CS) $Z=[z_1,\ldots,z_n]^T\in R^{n\times r}$ or diagnosis status (DS) (r=1 for Z)
Ensure: canonical vectors w_1,w_2,w_3
Initialization: $w_1\in R^{p\times 1},w_2\in R^{q\times 1},w_3\in R^{r\times 1}$
While not converged regarding to w_1,w_2,w_3 **do**
Calculate the diagonal matrix D_1, where the k_1-th element is $\frac{1}{2\,\|\,w_1^{k_1}\,\|_1}$;
Update $w_1=(X^TX+\beta_1D_1)^{-1}X^T(Yw_2+Zw_3)$;
Scale w_1 so that $\|Xw_1\|_2 = 1$;
Calculate the diagonal matrix D_2, where the k_2-th element is $\frac{1}{2\,\|\,w_2^{k_2}\,\|_1}$;
Update $w_2=(Y^TY+\beta_2D_2)^{-1}Y^T(Xw_1+Zw_3)$;
Scale w_2 so that $\|Yw_2\|_2 = 1$;
Calculate the diagonal matrix D_3, where the k_3-th element is $\frac{1}{2\,\|\,w_3^{k_3}\,\|_1}$;
Update $w_3=(Z^TZ+\beta_3D_3)^{-1}Z^T(Xw_1+Yw_2)$;
Scale w_3 so that $\|Zw_3\|_2 = 1$;
End while

References

1. Brookmeyer, R., Johnson, E., Ziegler-Graham, K. & Arrighi, H. M. Forecasting the global burden of Alzheimer's disease. *Alzheimers Dement* **3,** 186–191 (2007).
2. Pasinetti, G. M. & Hiller-Sturmhofel, S. Systems biology in the study of neurological disorders: focus on Alzheimer's disease. *Alcohol Res Health* **31,** 60–65 (2008).
3. Glahn, D. C., Thompson, P. M. & Blangero, J. Neuroimaging endophenotypes: Strategies for finding genes influencing brain structure and function. *Hum Brain Mapp* **28,** 488–501 (2007).
4. Gottesman, I. I. & Gould, T. D. The endophenotype concept in psychiatry: Etymology and strategic intentions. *Am J Psychiat* **160,** 636–645 (2003).
5. Tian Ge, G. S., Jianfeng Feng. Imaging genetics—towards discovery neuroscience. *Quant. Biol.* **1,** 227–245 (2013).
6. Shen, L. *et al.* Whole genome association study of brain-wide imaging phenotypes for identifying quantitative trait loci in MCI and AD: A study of the ADNI cohort. *NeuroImage* **53,** 1051–1063 (2010).
7. Stein, J. L. *et al.* Voxelwise genome-wide association study (vGWAS). *Neuroimage* **53,** 1160–1174 (2010).
8. Hibar, D. P., Kohannim, O., Stein, J. L., Chiang, M. C. & Thompson, P. M. Multilocus genetic analysis of brain images. *Front Genet* **2,** 73 (2011).
9. Kohannim, O. *et al.* Predicting Temporal Lobe Volume on Mri from Genotypes Using L(1)-L(2) Regularized Regression. *Proc IEEE Int Symp Biomed Imaging*, 1160–1163 (2012).
10. Wang, H. *et al.* Identifying quantitative trait loci via group-sparse multitask regression and feature selection: an imaging genetics study of the ADNI cohort. *Bioinformatics* **28,** 229–237 (2012).
11. Chi, E. C. *et al.* Imaging Genetics Via Sparse Canonical Correlation Analysis. *Proc IEEE Int Symp Biomed Imaging* **2013,** 740–743 (2013).
12. Witten, D. M., Tibshirani, R. & Hastie, T. A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. *Biostatistics* **10,** 515–534 (2009).
13. Witten, D. M. & Tibshirani, R. J. Extensions of Sparse Canonical Correlation Analysis with Applications to Genomic Data. *Stat Appl Genet Mol* **8** (2009).
14. Du, L. *et al.* Structured Sparse Canonical Correlation Analysis for Brain Imaging Genetics: An Improved GraphNet Method. *Bioinformatics* (2016).
15. Lin, D., Calhoun, V. D. & Wang, Y. P. Correspondence between fMRI and SNP data by group sparse canonical correlation analysis. *Med Image Anal* **18,** 891–902 (2014).
16. Chen, J., Bushman, F. D., Lewis, J. D., Wu, G. D. & Li, H. Structure-constrained sparse canonical correlation analysis with an application to microbiome data analysis. *Biostatistics* **14,** 244–258 (2013).
17. Chen, X. & Liu, H. An efficient optimization algorithm for structured sparse cca, with applications to eqtl mapping. *Statistics in Biosciences* **4,** 3–26 (2012).
18. Yan, J. *et al.* Transcriptome-guided amyloid imaging genetic analysis via a novel structured sparse learning algorithm. *Bioinformatics* **30,** i564–571 (2014).
19. Vounou, M., Nichols, T. E. & Montana, G. Discovering genetic associations with high-dimensional neuroimaging phenotypes: A sparse reduced-rank regression approach. *Neuroimage* **53,** 1147–1159 (2010).
20. Vounou, M. *et al.* Sparse reduced-rank regression detects genetic associations with voxel-wise longitudinal phenotypes in Alzheimer's disease. *Neuroimage* **60,** 700–716 (2012).
21. Batmanghelich, N. K., Dalca, A. V., Sabuncu, M. R. & Polina, G. Joint modeling of imaging and genetics. *Inf Process Med Imaging* **23,** 766–777 (2013).
22. Krishnan, A., Williams, L. J., McIntosh, A. R. & Abdi, H. Partial Least Squares (PLS) methods for neuroimaging: a tutorial and review. *Neuroimage* **56,** 455–475 (2011).
23. Le Cao, K. A., Rossouw, D., Robert-Granie, C. & Besse, P. A Sparse PLS for Variable Selection when Integrating Omics Data. *Stat Appl Genet Mol* **7** (2008).

24. Filippini, N. *et al.* Anatomically-distinct genetic associations of APOE epsilon4 allele load with regional cortical atrophy in Alzheimer's disease. *Neuroimage* **44,** 724–728 (2009).
25. Lambert, J. C. *et al.* Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet* **45,** 1452–1458 (2013).
26. Lin, R. *et al.* Association of common variants in TOMM40/APOE/APOC1 region with human longevity in a Chinese population. *J Hum Genet* **61,** 323–328 (2016).
27. Poulin, S. P. *et al.* Amygdala atrophy is prominent in early Alzheimer's disease and relates to symptom severity. *Psychiat Res-Neuroim* **194,** 7–13 (2011).
28. Risacher, S. L. *et al.* APOE effect on Alzheimer's disease biomarkers in older adults with significant memory concern. *Alzheimers Dement* **11,** 1417–1429 (2015).
29. Tuijl, J. P., Scholte, E. M., de Craen, A. J. & van der Mast, R. C. Screening for cognitive impairment in older general hospital patients: comparison of the Six-Item Cognitive Impairment Test with the Mini-Mental State Examination. *Int J Geriatr Psychiatry* **27,** 755–762 (2012).
30. Folstein, M. F., Folstein, S. E. & McHugh, P. R. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* **12,** 189–198 (1975).
31. Hughes, C. P., Berg, L., Danziger, W. L., Coben, L. A. & Martin, R. L. A new clinical scale for the staging of dementia. *Br J Psychiatry* **140,** 566–572 (1982).
32. Crane, P. K. *et al.* Development and assessment of a composite score for memory in the Alzheimer's Disease Neuroimaging Initiative (ADNI). *Brain Imaging Behav* **6,** 502–516 (2012).
33. Gibbons, L. E. *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* **6,** 517–527 (2012).
34. Kim, S. *et al.* Influence of genetic variation on plasma protein levels in older adults using a multi-analyte panel. *PLoS One* **8,** e70269 (2013).
35. Mahley, R. W. & Rall, S. C. Apolipoprotein E: Far more than a lipid transport protein. *Annu Rev Genom Hum G* **1,** 507–537 (2000).
36. Ashburner, J. & Friston, K. Voxel-Based Morphometry. *Statistical Parametric Mapping: The Analysis of Functional Brain Images*, 92–98 (2007).
37. Tzourio-Mazoyer, N. *et al.* Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* **15,** 273–289 (2002).

Acknowledgements

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.; 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 Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California. 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. This research is supported by the National Natural Science Foundation of China (Nos 61422204, 61473149,. 61501230), the NUAA Fundamental Research Funds (No. NE2013105), the Jiangsu Qinglan Project of China. At Indiana University, this work was supported by NIH R01 LM011360, U01 AG024904, RC2 AG036535, R01 AG19771, P30 AG10133, UL1 TR001108, R01 AG 042437, and R01 AG046171; NSF IIS-1117335; DOD W81XWH-14-2-0151, W81XWH-13-1-0259, and W81XWH-12-2-0012; NCAA 14132004; and CTSI SPARC Program.

Author Contributions

X.H., D.Z. and L.S. conceived and designed the research. X.H., C.L., L.D., and J.Y. carried out the study analysis. X.Y., S.L.R. and A.J.S. collected the data from ADNI database. X.H., D.Z., L.S. and A.J.S. analyzed the results and wrote the manuscript. Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report.

Additional Information

Competing Interests: The authors declare no competing financial interests.

How to cite this article: Hao, X. *et al.* Mining Outcome-relevant Brain Imaging Genetic Associations via Three-way Sparse Canonical Correlation Analysis in Alzheimer's Disease. *Sci. Rep.* **7**, 44272; doi: 10.1038/srep44272 (2017).

Consortia

Alzheimer's Disease Neuroimaging Initiative

Michael W. Weiner[5], Paul Aisen[6], Ronald Petersen[7,8], Clifford R. Jack Jr.[8], Sara S. Mason[8], Colleen S. Albers[8], David Knopman[8], Kris Johnson[8], William Jagust[9], John Q. Trojanowki[10], Arthur W. Toga[11], Laurel Beckett[12], Robert C. Green[13], Martin R. Farlow[13], Ann Marie Hake[13], Brandy R. Matthews[13], Jared R. Brosch[13], Scott Herring[13], Cynthia Hunt[13], Leslie M. Shaw[14], Beau Ances[14], John C. Morris[14], Maria Carroll[14], Mary L. Creech[14], Erin Franklin[14], Mark A. Mintun[14], Stacy Schneider[14], Angela Oliver[14], Jeffrey Kaye[15], Joseph Quinn[15], Lisa Silbert[15], Betty Lind[15], Raina Carter[15], Sara Dolen[15], Lon S. Schneider[11], Sonia Pawluczyk[11], Mauricio Beccera[11], Liberty Teodoro[11], Bryan M. Spann[11], James Brewer[16], Helen Vanderswag[16], Adam Fleisher[16,17], Pierre Tariot[17], Anna Burke[17], Nadira Trncic[17], Stephanie Reeder[17], Judith L. Heidebrink[18], Joanne L. Lord[18], Rachelle S. Doody[19], Javier Villanueva-Meyer[19], Munir Chowdhury[19], Susan Rountree[19], Mimi Dang[19], Yaakov Stern[20], Lawrence S. Honig[20], Karen L. Bell[20], Daniel Marson[21], Randall Griffith[21], David Clark[21], David Geldmacher[21], John Brockington[21], Erik Roberson[21], Marissa Natelson Love[21], Hillel Grossman[22], Effie Mitsis[22], Raj C. Shah[23], Leyla deToledo-Morrell[23], Ranjan Duara[24], Daniel Varon[24], Maria T. Greig[24], Peggy Roberts[24], Marilyn Albert[25], Chiadi Onyike[25], Daniel D'Agostino[25], Stephanie Kielb[25], James E. Galvin[26], Brittany Cerbone[26], Christina A. Michel[26], Dana M. Pogorelec[26], Henry Rusinek[26], Mony J. de Leon[26], Lidia Glodzik[26], Susan De Santi[26], P. Murali Doraiswamy[27], Jeffrey R. Petrella[27], Salvador Borges-Neto[27], Terence Z. Wong[27], Edward Coleman[27], Charles D. Smith[28], Greg Jicha[28], Peter Hardy[28], Partha Sinha[28], Elizabeth Oates[28], Gary Conrad[28], Anton P. Porsteinsson[29], Bonnie S. Goldstein[29], Kim Martin[29], Kelly M. Makino[29], M. Saleem Ismail[29], Connie Brand[29], Ruth A. Mulnard[30], Gaby Thai[30], Catherine Mc-Adams-Ortiz[30], Kyle Womack[31], Dana Mathews[31], Mary Quiceno[31], Allan I. Levey[32], James J. Lah[32], Janet S. Cellar[32], Jeffrey M. Burns[33], Russell H. Swerdlow[33], William M. Brooks[33], Liana Apostolova[34], Kathleen Tingus[34], Ellen Woo[34], Daniel H. S. Silverman[34], Po H. Lu[34], George Bartzokis[34], Neill R. Graff-Radford[35], Francine Parfitt[35], Tracy Kendall[35], Heather Johnson[35], Christopher H. van Dyck[36], Richard E. Carson[36], Martha G. MacAvoy[36], Pradeep Varma[36], Howard Chertkow[37], Howard Bergman[37], Chris Hosein[37], Sandra Black[38], Bojana Stefanovic[38], Curtis Caldwell[38], Ging-Yuek Robin Hsiung[39], Howard Feldman[39], Benita Mudge[39], Michele Assaly[39], Elizabeth Finger[40], Stephen Pasternack[40], Irina Rachisky[40], Dick Trost[40], Andrew Kertesz[40], Charles Bernick[41], Donna Munic[41], Marek- Marsel Mesulam[42], Kristine Lipowski[42], Sandra Weintraub[42], Borna Bonakdarpour[42], Diana Kerwin[42], Chuang-Kuo Wu[42], Nancy Johnson[42], Carl Sadowsky[43], Teresa Villena[43], Raymond Scott Turner[44], Kathleen Johnson[44], Brigid Reynolds[44], Reisa A. Sperling[45], Keith A. Johnson[45], Gad Marshall[45], Jerome Yesavage[46], Joy L. Taylor[46], Barton Lane[46], Allyson Rosen[46], Jared Tinklenberg[47], Marwan N. Sabbagh[47], Christine M. Belden[47], Sandra A. Jacobson[47], Sherye A. Sirrel[47], Neil Kowall[48], Ronald Killiany[48], Andrew E. Budson[48], Alexander Norbash[48], Patricia Lynn Johnson[48], Thomas O. Obisesan[49], Saba Wolday[49], Joanne Allard[49], Alan Lerner[50], Paula Ogrocki[50], Curtis Tatsuoka[50], Parianne Fatica[50], Evan Fletcher[51], Pauline Maillard[51], John Olichney[51], Charles DeCarli[51], Owen Carmichael[51], Smita Kittur[52], Michael Borrie[53], T.-Y. Lee[53], Rob Bartha[53], Sterling Johnson[54], Sanjay Asthana[54], Cynthia M. Carlsson[54], Steven G. Potkin[55], Adrian Preda[55], Dana Nguyen[55], Vernice Bates[56], Horacio Capote[56], Michelle Rainka[56], Douglas W. Scharre[57], Maria Kataki[57], Anahita Adeli[57], Earl A. Zimmerman[58], Dzintra Celmins[58], Alice D. Brown[58], Godfrey D. Pearlson[59], Karen Blank[59], Karen Anderson[59], Laura A. Flashman[60], Marc Seltzer[60], Mary L. Hynes[60], Robert B. Santulli[60], Kaycee M. Sink[61], Leslie Gordineer[61], Jeff D. Williamson[61], Pradeep Garg[61], Franklin Watkins[61], Brian R. Ott[62], Henry Querfurth[62], Geoffrey Tremont[62], Stephen Salloway[63], Paul Malloy[63], Stephen Correia[63], Howard J. Rosen[64], Bruce L. Miller[64], David Perry[64], Jacobo Mintzer[65], Kenneth Spicer[65], David Bachman[65], Nunzio Pomara[66], Raymundo Hernando[66], Antero Sarrael[66], Norman Relkin[67], Gloria Chaing[67], Michael Lin[67], Lisa Ravdin[67], Amanda Smith[68], Balebail Ashok Raj[68] & Kristin Fargher[68]

[5]Magnetic Resonance Unit at the VA Medical Center and Radiology, Medicine, Psychiatry and Neurology, University of California, San Francisco, USA. [6]San Diego School of Medicine, University of California, California, USA. [7]Mayo Clinic, Minnesota, USA. [8]Mayo Clinic, Rochester, USA. [9]University of California, Berkeley, USA. [10]University of Pennsylvania, Pennsylvania, USA. [11]University of Southern California, California, USA. [12]University of California, Davis, California, USA. [13]MPH Brigham and Women's Hospital/Harvard Medical School, Massachusetts, USA.

[14]Washington University St. Louis, Missouri, USA. [15]Oregon Health and Science University, Oregon, USA. [16]University of California–San Diego, California, USA. [17]Banner Alzheimer's Institute, USA. [18]University of Michigan, Michigan, USA. [19]Baylor College of Medicine, Houston, State of Texas, USA. [20]Columbia University Medical Center, South Carolina, USA. [21]University of Alabama – Birmingham, Alabama, USA. [22]Mount Sinai School of Medicine, New York, USA. [23]Rush University Medical Center, Rush University, Illinois, USA. [24]Wien Center, Florida, USA. [25]Johns Hopkins University, Maryland, USA. [26]New York University, NY, USA. [27]Duke University Medical Center, North Carolina, USA. [28]University of Kentucky, Kentucky, USA. [29]University of Rochester Medical Center, NY, USA. [30]University of California, Irvine, California, USA. [31]University of Texas Southwestern Medical School, Texas, USA. [32]Emory University, Georgia, USA. [33]University of Kansas, Medical Center, Kansas, USA. [34]University of California, Los Angeles, California, USA. [35]Mayo Clinic, Jacksonville, USA. [36]Yale University School of Medicine, Connecticut, USA. [37]McGill University, Montreal-Jewish General Hospital, Canada. [38]Sunnybrook Health Sciences, Ontario, USA. [39]U.B.C. Clinic for AD & Related Disorders, Canada. [40]Cognitive Neurology - St. Joseph's, Ontario, USA. [41]Cleveland Clinic Lou Ruvo Center for Brain Health, Ohio, USA. [42]Northwestern University, USA. [43]Premiere Research Inst (Palm Beach Neurology), USA. [44]Georgetown University Medical Center, Washington D.C, USA. [45]Brigham and Women's Hospital, Massachusetts, USA. [46]Stanford University, California, USA. [47]Banner Sun Health Research Institute, USA. [48]Boston University, Massachusetts, USA. [49]Howard University, Washington D.C, USA. [50]Case Western Reserve University, Ohio, USA. [51]University of California, Davis – Sacramento, California, USA. [52]Neurological Care of CNY, USA. [53]Parkwood Hospital, Pennsylvania, USA. [54]University of Wisconsin, Wisconsin, USA. [55]University of California, Irvine – BIC, USA. [56]Dent Neurologic Institute, NY, USA. [57]Ohio State University, Ohio, USA. [58]Albany Medical College, NY, USA. [59]Hartford Hospital, Olin Neuropsychiatry Research Center, Connecticut, USA. [60]Dartmouth-Hitchcock Medical Center, New Hampshire, USA. [61]Wake Forest University Health Sciences, North Carolina, USA. [62]Rhode Island Hospital, state of Rhode Island, USA. [63]Butler Hospital, Providence, Rhode Island, USA. [64]University of California, San Francisco, USA. [65]Medical University South Carolina, USA. [66]Nathan Kline Institute, Orangeburg, New York, USA. [67]Cornell University, Ithaca, New York, USA. [68]USF Health Byrd Alzheimer's Institute, University of South Florida, USA.