Susceptibility of brain atrophy to TRIB3 in Alzheimer’s disease: evidence from functional prioritization in imaging-genetics

Marco Lorenzi1, Andre Altmann2, Boris Gutman3, Selina Wray4, Charles Arber4, Derrek Hibar5, Neda Jahanshad3, Jonathan Schott1, Daniel Alexander1, Paul Thompson1, Sebastien Ourselin6

1Université Côte d’Azur, 2Translational Imaging Group, Centre for Medical Image Computing (CMIC), Department of Medical Physics & Bioengineering, University College London, London WC1E 6BT, United Kingdom, 3University of Southern California, 4University College London, 5USC, 6Institute of Neurology, UCL

Submitted to Proceedings of the National Academy of Sciences of the United States of America

The joint modelling of brain imaging information and genetic data is a promising research avenue to highlight the functional role of genes in determining the pathophysiological mechanisms of Alzheimer’s disease (AD). However, since genome-wide association (GWA) studies are essentially limited to the exploration of statistical correlations between genetic variants and phenotype, the validation and interpretation of the findings is usually non trivial and prone to false positives. To address this issue, in this work we investigate the genetic functional mechanisms underlying brain atrophy in AD by studying the involvement of candidate variants in known genetic regulatory functions. This approach, here termed functional prioritization, aims at testing the sets of gene-variants identified by high-dimensional multivariate statistical modelling with respect to known biological processes, in order to introduce a biology-driven validation scheme. When applied to the ADNI cohort, the functional prioritization allowed identifying a link between TRIB3 (tribbles pseudokinase 3) and the stereotypical pattern of grey matter loss in AD, which was confirmed in an independent validation sample, and that provides novel evidence about the relation between this gene and known mechanisms of neurodegeneration.

Introduction

Alzheimer’s disease (AD) is a devastating neurodegenerative disorder and its aetiology still remains largely concealed. In the anticipation of increasing prevalence of AD and other dementias, there is an urgent need for improving the understanding of the disease processes that underlie neurodegeneration. Whilst the knowledge about the genetic and environmental risks underpinning AD is steadily advancing, our understanding of how these factors interact to lead to the complex pathophysiology that results in dementia is less understood.

Advances in imaging technologies have led to non- or minimally-invasive imaging biomarkers that capture various aspects of the disease process including amyloid deposition [1], tau pathology [2], functional decline [3] and neuronal loss [4]. Combining such imaging information with genetic measurements – so called imaging-genetics – provides the means for investigating the effect of genetic variation on underlying biological mechanisms [5].

Genome-wide association studies (GWAS) query millions of single nucleotide polymorphisms (SNPs) individually for their association with either case-control status [6] or disease-specific quantitative phenotypes, e.g., in the case of AD, regional brain volumes [7] or brain amyloid burden [8]. Mass univariate analysis of genetic data is still the predominant method, in virtue of its ease-of-use and well-established theoretical framework, albeit suffering from significant limitations including the requirement for multiple testing, redundancies introduced by linkage disequilibrium (LD) and the lack of analysis of epistatic effects (e.g., SNP-SNP interactions), which have to be explicitly modeled and searched for exhaustively [9]. Moreover, more than one quantitative phenotype can be derived from the available imaging data, e.g., dozens or hundreds of regional brain volumes, or hundreds of thousands of voxel-level metrics [10]. This potentially large number of genotype-phenotypes features of interest generally complicates the problem of reliably detecting statistical associations, and thus hampers the identification of disease-relevant genetic markers by purely statistical means.

Limitations of classical mass-univariate statistical methods have in recent years been overcome by employing multivariate approaches to data analysis in the context of neuroscience studies [11] and GWAS [12]. Likewise, in imaging-genetics meaningful genotype-phenotype interactions [13] are captured by simultaneously modeling sets of genetic variants that are jointly associated with a given imaging phenotype [14,15,16,17]. Multivariate GWAS have the potential to shed light on the complex genotype-phenotype relationship, and may thus highlight novel links between brain physiology and molecular and biological functions. However, although these methods have proven their ability to

Significance

In this study we employ a novel experimental imaging-genetics approach for investigating the genetic underpinnings of brain atrophy in Alzheimer’s disease. We successfully combined state-of-art imaging-genetics methods and experimental gene expression data to uncover novel biology in brain atrophy. The novel experimental paradigm highlighted a significant role of TRIB3 (tribbles pseudokinase 3) in modulating the typical pattern of Alzheimer’s brain pathology. This result corroborates through rigorous data-driven statistical methods evidence emerging from previous studies about the role of TRIB3 in modulating known mechanisms of neurodegeneration, such as neuronal death, cellular homeostasis, and interaction with established genes causing autosomal dominant Alzheimer's disease: APP and PSEN1. The developed integrated statistical-experimental methodology could serve as a roadmap for investigations in other disorders.

Reserved for Publication Footnotes
findings remain very challenging tasks. These problems relate directly to the understanding of the functional role of sets of genetic variants, and to the difficulty of replicating the statistical results in unseen cohorts.

We approach this technical bottleneck by leveraging multivariate approaches to explore high-dimensional datasets and to generate hypotheses, which are subsequently tested in downstream experiments. High-quality databases of matched genotype and gene expression measurements such as the Genotype-Tissue-Expression project (GTEx) [18] and BRAINEAC [19] facilitate the quantification of effects of SNPs on gene expression in numerous tissues, including various brain tissues. Typically, these databases are used to detail the effect of a genetic variant at the very end of an analysis pipeline and to garner evidence for molecular mechanisms of the genetic loci. However, functional information in ‘convenience’ databases can also be used at an earlier stage in the analysis in order to prioritize a few candidate hypotheses with a clear functional mechanism (e.g., expression quantitative trait loci; eQTL) for the validation phase and thus limit the multiple testing burden.

In this work we apply this novel investigative approach to study the genetic functional mechanisms underlying brain atrophy in AD. The framework is comprised of two steps:

i) **Statistical discovery.** Candidate genetic variants are initially identified through data-driven multivariate statistical analysis of the matched imaging and genetics data (Figure 1). This is achieved by modeling the joint covariance between 1.1 million SNPs and the cortical and subcortical atrophy represented by 327,684 cortical and 27,120 sub-cortical thickness values of 639 individuals (either healthy older controls or patients with AD) from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) cohort.

ii) **Functional prioritization.** The candidate genetic variants are subsequently screened for functional relevance by querying high-dimensional gene expression databases such as GTEx.

The resulting small set of genetic loci, which are shown to modify gene expression, is then validated in an independent sample of 553 individuals from ADNI diagnosed with mild cognitive impairment (MCI), a proportion of whom progressed to AD.

Compared to previous approaches our work (i) analyses the whole genome on the whole brain in a hypothesis free fashion,
Table 1. Socio-demographic, clinical, and genetic characteristic of the study cohort (mean, standard deviation). MCi: MCi individual subsequently converted to Alzheimer’s disease. MCIc: MCI individual not converted to Alzheimer’s disease during the observational time. MMSE: mini mental state examination. ADAS: Alzheimer’s Disease Assessment Scale, 11 items. ApoE: Apolipoprotein E, allele 4. Positivity to Aβ1-42 was defined with respect to the nominal cut-off of 192pg/ml.

Table 2. Statistical comparison of the genes scores in training and testing groups (Kruskal Wallis nonparametric test). The score for TRIB3 leads to significant differences in the MCI testing group after Bonferroni correction for multiple comparisons.

<table>
<thead>
<tr>
<th>gene</th>
<th>p-value training (AD vs CT)</th>
<th>p-value testing (MCI conv vs MCI stable)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TM2D1</td>
<td>0.0050</td>
<td>0.0528</td>
</tr>
<tr>
<td>IL10RA</td>
<td>0.0169</td>
<td>0.6198</td>
</tr>
<tr>
<td>TRIB3</td>
<td>0.0032</td>
<td>0.0034</td>
</tr>
<tr>
<td>ZBTB7A</td>
<td>0.0360</td>
<td>0.9135</td>
</tr>
<tr>
<td>LYSMD4</td>
<td>0.0000</td>
<td>0.2057</td>
</tr>
<tr>
<td>CRY1L</td>
<td>0.6213</td>
<td>0.1176</td>
</tr>
<tr>
<td>FAM135B</td>
<td>0.0000</td>
<td>0.5588</td>
</tr>
<tr>
<td>IP6K3</td>
<td>0.0000</td>
<td>0.4646</td>
</tr>
<tr>
<td>ITGA1</td>
<td>0.0993</td>
<td>0.7310</td>
</tr>
<tr>
<td>KIN</td>
<td>0.0014</td>
<td>0.2061</td>
</tr>
<tr>
<td>LAMC1</td>
<td>0.0019</td>
<td>0.0618</td>
</tr>
<tr>
<td>LINC00941</td>
<td>0.0000</td>
<td>0.6896</td>
</tr>
<tr>
<td>RBPMS2</td>
<td>0.0000</td>
<td>0.2149</td>
</tr>
<tr>
<td>RP11-181K3.4</td>
<td>0.0017</td>
<td>0.0527</td>
</tr>
</tbody>
</table>

i.e., without preselecting SNPs or brain regions and (ii) uses a functional prioritization step in order to select genetic loci for validation in an independent cohort. Starting from the initial ~1.1 million SNPs, the multivariate statistical analysis allowed the identification of a relatively small number of genetic loci that are statistically associated with the typical pattern of AD brain pathology. The subsequent functional prioritization step ultimately identified a significant role of TRIB3 (tribles pseudokinase 3), a gene showing important connections to known mechanisms of neurodegenerative diseases. Indeed, although a role for TRIB3 in dementia has not been extensively explored, there are several aspects of TRIB3 function that have relevance to mechanisms related to neuronal death, cellular homeostasis, and of interaction with established AD genes, such as APP and PSEN1.

This study ultimately offers an illustration of the potential of effectively combining multivariate statistical modeling in imaging-genetics with recent instruments available from computational biology, to lead to novel insights on the pathophysiology of neurodegeneration.

Results

Model training and estimated components

Figure 2 and 3 show the relevant areas of the identified joint genetic and phenotype variation, respectively, for the first three PLS components through stability selection. The components were very robust (100% reproducible) during the stability selection procedure (Supplementary Methods). The fourth and fifth components did not present any relevant locations (i.e., all bins have p<0.05) after stability selection for both the genetic modality and for the imaging modality.

Genetic components

The circular Manhattan plot (Circos v0.96 [20]) of Figure 2 shows the selection frequency for the PLS genotype components, describing the importance of the genetic loci associated to cortical thickness variation for component 1, 2 and 3. The plot shows the probability of a given genetic bin of size 10kb of being relevant in the PLS model, i.e., to contain a SNP that is ranked in the top 10% of the absolute weights of the genotype component. Spatially contiguous loci generally show similar importance values, which is caused by LD of these regions.

In the genetic components 1 through 3 a total of 118 bins exceeded the selection frequency threshold (61, 50, 7 for component 1, 2 and 3, respectively). From these bins 402 (196, 181 and 25) influential SNPs were extracted and annotated with 98 genes through the Ensembl Variant Effect Predictor (VEP) for GRCh37 (date accessed: 17th October 2016) [21]. The extended APOE locus comprising APOE and TOMM40 was selected as the highest scoring region in component 1. A total of 3,956 candidate SNP-gene pairs were considered for the GTEX-based eQTL analysis in six tissues. However, a few genes did not show sufficient expression levels in some tissues and these combinations were excluded from the analysis, resulting in 1,598 unique SNP-gene-tissue tests, of those 104 were significant at the Bonferroni corrected p-value threshold (p=3.1e-5) (Dataset S1) linking to 14 genes (Dataset S2; Figure S5): TRIB3, APOE, rs62191440, IP6K3, ITGA1, KIN, LAMC1, LINC00941, LYSMD4, RBPMS2, RP11-181K3.4, TM2D1, and TRIB3. These genes are listed in the innermost circle of Figure 2 depending on their genomic position.

The independent validation of those 14 genes in the MCI cohort confirmed TRIB3 (p=0.0034) (Table 2). Three additional genes were close to nominal significance: TM2D1 (p=0.053), LAMC1 (p=0.062), and RP11-181K3.4 (p=0.053) (Table 2). Of note the top eQTL SNP for TRIB3 rs4813620 received a p=0.06175 in stage I of a large AD GWAS [6]. However, rs62191440, a SNP in strong LD with rs4813620 (D=0.8469; r²=0.6559) in the European population [22], received a p-value of 0.00601 (Figure S6) and also constitutes an eQTL for TRIB3 in various tissues in GTEX including brain tissues cortex and caudate ganglia (Figure S7). Interestingly, when estimating the PLS com-
In this work we modeled high-dimensional genome-wide SNP data and brain-wide cortical thickness data via joint multivariate statistical modeling and functional prioritization of genes through bioinformatics annotation and a large eQTL database. The absolute value of the weights is proportional to the importance of the underlying brain areas. The relevance of the brain areas is quantified in the bottom row. The colors (red to white) indicate the probability of a brain area to be associated with the genotype component shown in Figure 2, and quantify the probability of each cortical mesh points of being relevant in the PLS model, i.e., to be ranked among the top 10% of the absolute weights of the phenotype component.

Morphometric components

Figure 3 shows the PLS phenotype components 1 through 3 along with the associated selection frequency describing the loci of brain atrophy associated with genetic variation. The first component is mainly associated to the thinning of the cortical mantle, and is localized in temporal and posterior cingulate cortices. The relevant areas at the subcortical level are primarily associated with amygdalae and thalami. The second component is mostly associated to the thinning of the subcortical areas (hippocampi and amygdalae), and to the cortical thinning of the temporal areas at the cortical level. The third component is similar to component 2, and describes a sub-cortical thickness pattern prevalent in hippocampi, amygdalae, and thalami. At the cortical level, the component is associated with the thinning of frontal cortices, and to isolated spots located in the parahippocampal gyrus.

Discussion

In this work we modeled high-dimensional genome-wide SNP data and brain-wide cortical thickness data via joint multivariate statistical modeling and functional prioritization of genes through bioinformatics annotation and a large eQTL database.

Our study ultimately identified a link between TRIB3 (troubles pseudokinase 3) and the stereotypical pattern of grey matter loss in AD (cortical thinning in temporal and posterior cingulate regions and subcortical atrophy). TRIB3 is a pseudokinase which acts as a regulator of several signaling pathways. For example it can interact directly with Akt and inhibit the pro-survival Akt pathway [23]. TRIB3 expression is induced during neuronal cell death [24] and recently increased levels of the TRIB3 protein were found in dopaminergic neurons of the substantia nigra pars compacta in patients with Parkinson’s disease [25]. TRIB3 expression is stress induced and increases in response to nerve growth factor (NGF) deprivation; endoplasmic reticulum (ER) stress, and amino acid deprivation [24]. Although a role for TRIB3 in dementia has not been extensively explored, there are several aspects of TRIB3 function that have relevance to known mechanisms of neurodegenerative disease. TRIB3 can interact directly with P62 to modulate autophagic flux [26], an important process in maintaining cellular homeostasis that is known to be disrupted in neurodegeneration [27]. Knockdown of TRIB3 modulates PSEN1 stability [26] and a yeast two-hybrid screen identified progranulin as a direct interaction partner of TRIB3 [28]. Intriguingly, it has recently been demonstrated that TRIB3 induces both apoptosis and autophagy in Aβ-induced neuronal death, and silencing of TRIB3 was strongly neuroprotective [29]. These links warrant further investigation for a functional role of TRIB3 in neuronal death in dementia.

These earlier findings align with our eQTL analysis where carriers of the minor allele show increased TRIB3 expression (Figure S5), which potentially lowers the threshold to TRIB3 mediated neuronal cell death. TRIB3 expression was modulated by the identified SNP in various other tissues including the caudate (Figure S7), a region affected in PD and Huntington’s disease. A recent study of Trib3 expression in mice concluded that “Trib3 has a pathophysiological role in diabetes” [30]; diabetes itself is a known risk factor for dementia [31] perhaps through shared metabolic processes with AD [32]. Interestingly, one of the three SNPs (rs1555318) selected in the PLS model and attributed to TRIB3 showed a strong association with type-2 diabetes in stage 1 of a large GWAS (p=4.4e-4; Figure S8) [33]. Other GWAS showed links between TRIB3 and information processing speed (p=1.7e-7) [34] and AD (p=0.006; [6]). An earlier genetic study on AD in Swedish men found an association in TRIB3 as
The functional prioritization component of the analysis successfully reduced the set of candidate genetic variants for the independent validation, however, this prioritization has a shortcoming: it hypothesizes that identified SNPs alter the expression of a nearby gene. Although, this scheme led to the identification of TRIB3 in the cortical thickness phenotype, it did miss a long-established AD risk gene: APOE. SNPs belonging to APOE (rs429358 and rs7412) were selected as highest scoring SNPs in component 1. However, none of them was detected as an eQTL and thus APOE was excluded from the downstream analysis. Other types of functional prioritizations based on exonic function prediction may have retained APOE and other genes in the pipeline. However, SNPs data typically features only a few non-synonymous exonic variants and their high frequency (MAF > 5%) renders them unlikely to receive significant damaging scores in these predictions. Thus, for this scenario the use of these function predictions would be limited.

The list of genes we identified contains other interesting candidates. For instance, IL10RA (interleukin 10 receptor subunit alpha) is a receptor for interleukin 10 (IL10), a cytokine that controls inflammatory response [38]. Carriers of the minor allele showed increased IL10RA expression (Figure S5) and IL10ra expression is increased in affected brain regions with increasing age and presence of AD pathology in transgenic mouse models of AD (MOUSEAC; [39] Figure S9). Moreover, a link between downregulation of IL10RA and TRIB3 in TRIB3-silenced HepG2 cells was reported in [26], along with increased abundance of Pre-selinin 1, ApoE3, and Clusterin. Finally, blocking IL10 response was recently suggested as a therapeutic mechanism in AD [40].

A gene that showed a statistical trend in the validation sample was TM2D1 (TM2 domain containing 1), which is a beta-amyloid binding protein and may be involved in beta-amyloid-induced apoptosis [41]. Further, MEF2A (Myocyte Enhancer Factor 2A), like APOE, was filtered out by the functional prioritization. However, MEF2A is a paralog of MEF2C, which is an established AD gene [6]. Noteworthy, bins covering MEF2C only barely missed the selection threshold in component 2 for further analysis (max p=0.926; Figure 2).

This study illustrates the potential of effectively combining multivariate statistical modeling in imaging-genetics with recent instruments available from computational biology, to lead to novel insights regarding lipid metabolism. Thanks to the ever-growing data-driven knowledge based on the vast quantities of information now available to the research community, the paradigm proposed in this study may represent a promising avenue for linking imaging-genetics findings to the current knowledge on functional genetics mechanisms involved in neurodegeneration.

Materials and methods

This section describes the study data, the statistical setting and methodology used in the present study. Further details and discussion about the methodological aspects can be found in the Supplementary Material (Methods Section).

Study Participants

Data used in the preparation of this article were obtained from the ADNI database (http://adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. For up-to-date information, see www.adni-info.org. This research mainly involves further processing of previously collected personal data. We have explicit authorization for the use of the ADNI dataset, and we have signed the relevant papers guaranteeing that we abided to the ethics standards.

We selected genotype and phenotype data available in the ADNI-1QG02 datasets for 1,192 subjects. Summary socio-demographic, clinical and genetic information are available in Table 1. At time of study entry subjects were diagnosed as healthy individuals (N=401), MCI (N=553) or AD (N=238). A total of 212 (38.3%) MCI patients subsequently converted to AD over the course of the study (6 years). All participants were non-Hispanic Caucasian. AD and MCI groups show significant cognitive decline measured by MMSE and ADAS-COG as compared to the healthy individuals (p<1e-2, two sample t-test for group-wise comparison). There was also a significant increase of individuals with pathological values of AB42 in the CSF (Aβ1-42 >192pg/ml) across the disease stages, with proportion ranging from 43% for healthy individuals to 93% for AD patients (p<1e-2). Similarly, we observed a higher prevalence of APOE4 carriers in AD and progressing MCI individuals when compared to healthy and MCI stable individuals. For this analysis, the 639 healthy and AD subjects form the discovery set, while the MCI converters and non-converters form the independent validation set.

Statistical Discovery

The joint relationship between the genetic and imaging modalities was investigated through partial least squares (PLS) modeling [42,43,44]. Among the several PLS versions proposed in the literature we focus on the symmetric formulation of PLS computed through the singular value decomposition (SVD) of the cross covariance matrix (Figure S1) [43,45,46]. Within this setting, the aim of PLS is to estimate the latent components that maximize the global weight in the latent component that represents its relative importance for describing the global joint multimodal relationship. Analyzing these weights helps identifying SNPs that are linked to the patterns of cortical thinning in the brain.

In this study we applied a robust approach for the stable estimation and interpretation of PLS weights in genome-wide genotyping data, aimed at promoting sparsity (i.e., selecting only few features for simplified interpretation) and regularity (by aggregating SNPs within the same genetic neighborhood). This is achieved through a stability selection procedure in which the reproducibility of the PLS components is assessed through a split-half cross-validation based scheme on 1,000,000 repetitions of the models on randomly sampled subgroups (Figure 1 and Supplementary Methods). By considering a pre-defined partition of each chromosome into contiguous loci of size 10kb, the reproducibility of the first five PLS modes was calculated through a half procedure, and to assess the significance of the models, the average reproducibility of the first five PLS modes was calculated. Furthermore, we performed analysis of the top PLS models for each variant, and the results were used to identify the most discriminative features.

Gene identification

We analyzed the 10kb bins (genetic loci) with the selection frequency exceeding 0.95, i.e., bins selected in 95% or more of the 1,000,000 replications. Within these bins we then identified the influential SNPs: a SNP was declared influential if it was associated with the weights of highest magnitude in the PLS components estimated on the full data sample, i.e., SNPs with absolute weights exceeding the 99% quantile of all weights in the component. For this analysis, the 639 healthy and AD subjects form the discovery set, while the MCI converters and non-converters form the independent validation set.

In order to link SNPs to corresponding genes we used the computational VEP for GRCh37 with the GENCODE gene annotation. SNPs tagged as ‘regulatory’ were manually investigated and annotated with the nearby gene.

Functional prioritization

All SNPs successfully annotated with a gene were subjected to functional prioritization through expression quantitative trait loci (eQTL) analysis based on the GTEx data. The sample size in GTEx for relevant brain tissues in AD was rather small (e.g., N=81 for hippocampus). Therefore, we added five more tissues with large sample sizes that were more distantly relevant to AD: nerve tissue (N=256) was added as a proxy for nervous tissue; whole blood (N=338) and artery (N=285) were included to cover blood-based changes and effects on blood vessels (47) adipose subcutaneous (N=288) was selected due to links between AD and obesity, type-2 diabetes and metabolic disease [48,49]. Finally, transformed fibroblasts (N=272) were included as a general-purpose cell line. P-values were corrected for multiple testing using the Bonferroni correction.

Model validation in independent MCI subjects

The genes that were found to be under expression control by the identified SNPs were validated for their capacity to predict clinical conversion in MCI subjects. To this end, for each identified gene we applied the PLS weights estimated on the discovery set on the validation set, with the genetic component restricted to SNPs +/- 20kb of the gene borders. The identified latent projection (i.e., a weighted combination of the scores of one SNP per subject) for each gene along with the associated p-value were computed. The association between the projection score and conversion status was assessed by statistically comparing the scores distribution between healthy individuals and AD patients, and between MCI converters and non-converters (Kruskal-Wallis non parametric test for two sample comparison, Bonferroni correction for multiple comparisons).

Acknowledgment

Footnote Author
SO receives funding from the EPSRC (EP/H046410/1, EP/K005278), the MRC (MR/L023784/1), the PTP project VPH-DAREBIT (FP7-ICT-2011-9-501-055), the NIHR Biomedical Research Unit (Dementia) at UCL and the National Institute for Health Research University College London Hospitals Biomedical Research Centre (NIHR BRC UCL/UCL High Impact Initiative-BW:mm.BRC10-268). EPSRC grants EP/O029990/1 and EP/W02533/1 support DA and SO's work on this topic. ML, DA, JS, and SO also received support from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 634005 (MS acknowledges the support of the National Institute for Health Research University College London Hospitals Biomedical Research Centre, Wolfson Foundation, EPSRC (EP/L020990/1), ARUK (ARUK-Network2012-6-ICE; ARUK-PG2017-1946; ARUK-PG2017-1946), Brain Research Trust (UCC14191) and European Union's Horizon 2020 research and innovation programme (Grant 665992). AA holds an MRC eMedLab Medical Bioinformatics Career Development Fellowship. This work was supported by the Medical Research Council (G0880111). SW and CA are supported by the NIHR Queen Square Biomedical Research Unit in Dementia and Alzheimer's Research UK, and by the National Institute for Health Research University College London Hospitals Biomedical Research Centre. Contribution to this work by BG and PT was funded by the National Institutes of Health "Data to Knowledge" (BD2K) award (NIH U54 EB020403, P1. Thompson). Data collection and sharing for this project was funded by the Alzheimer’s Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024944) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). The detailed acknowledgement list for ADNI is provided in the supplementary material. Conflict of Interest The authors do not report a financial or personal relationship with any company or entity. Footnotes 1 http://gtexportal.org/home/ 2 http://www.brainecog.org/