

RESEARCH

Open Access



Identifying Alzheimer's genes via brain transcriptome mapping

Jae Young Baik^{1†}, Mansu Kim^{2†}, Jingxuan Bao³, Qi Long³, Li Shen^{3*}  Alzheimer's Disease Neuroimaging Initiative

From International Conference on Intelligent Biology and Medicine (ICIBM 2021) Philadelphia, PA, USA. 8-10 August 2021

Abstract

Background: Alzheimer's disease (AD) is one of the most common neurodegenerative disorders characterized by progressive decline in cognitive function. Targeted genetic analyses, genome-wide association studies, and imaging genetic analyses have been performed to detect AD risk and protective genes and have successfully identified dozens of AD susceptibility loci. Recently, brain imaging transcriptomics analyses have also been conducted to investigate the relationship between neuroimaging traits and gene expression measures to identify interesting gene-traits associations. These imaging transcriptomic studies typically do not involve the disease outcome in the analysis, and thus the identified brain or transcriptomic markers may not be related or specific to the disease outcome.

Results: We propose an innovative two-stage approach to identify genes whose expression profiles are related to diagnosis phenotype via brain transcriptome mapping. Specifically, we first map the effects of a diagnosis phenotype onto imaging traits across the brain using a linear regression model. Then, the gene-diagnosis association is assessed by spatially correlating the brain transcriptome map with the diagnostic effect map on the brain-wide imaging traits. To demonstrate the promise of our approach, we apply it to the integrative analysis of the brain transcriptome data from the Allen Human Brain Atlas (AHBA) and the amyloid imaging data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) cohort. Our method identifies 12 genes whose brain-wide transcriptome patterns are highly correlated with six different diagnostic effect maps on the amyloid imaging traits. These 12 genes include four confirmatory findings (i.e., AD genes reported in DisGeNET) and eight novel genes that have not been associated with AD in DisGeNET.

Conclusion: We have proposed a novel disease-related brain transcriptomic mapping method to identify genes whose expression profiles spatially correlated with regional diagnostic effects on a studied brain trait. Our empirical study on the AHBA and ADNI data shows the promise of the approach, and the resulting AD gene discoveries provide valuable information for better understanding biological pathways from transcriptomic signatures to intermediate brain traits and to phenotypic disease outcomes.

[†]Jae Young Baik and Mansu Kim have contributed equally

*Correspondence: li.shen@pennmedicine.upenn.edu

³ Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, USA
Full list of author information is available at the end of the article



Keywords: Brain imaging transcriptomics, Imaging-diagnosis map, Gene expression map

Background

Alzheimer’s disease (AD) is one of the most common neurodegenerative disorders characterized by a decline in cognitive function. Given that many studies demonstrated the high heritability of AD, ranging from 60 to 80%, researchers have performed targeted genetic studies or genome-wide association studies (GWAS) to link genetic variants such as single nucleotide polymorphisms (SNPs) to clinical outcomes [1].

These conventional genetic association studies discovered many AD-related genetic variants, including those from genes such as apolipoprotein E (APOE), clusterin (CLU), ATP-binding cassette, sub-family A (ABC1), member7 (ABCA7), complement component(3b/4b) receptor1 (CR1), and others [2–4]. However, it remains a challenge to interpret and validate these SNP-diagnosis associations directly since there is a lack of intermediate molecular signatures (e.g., epigenomic, transcriptomic, proteomic, and metabolomic measures) and system-level biomarkers (fluid biomarkers, brain structural and functional measures, and cognitive and behavioral measures) to link genetics to disease status [5–7].

Recent studies have investigated the associations between molecular features (e.g., gene expression) with disease-related imaging traits [8–11]. The assays of gene expression, in particular, offer a unique opportunity to measure gene functions. The Allen Human Brain Atlas (AHBA) provides us comprehensive whole-genome whole-brain microarray transcriptomics data. Many researchers have documented molecular and cellular processes as well as the brain structural and functional

changes involved in neurodegeneration by measuring spatial correlations between gene expression patterns and brain imaging phenotypes. For example, prior studies employed a correlation or regression model to link neuroimaging features (e.g., transverse relaxation time and regional atrophy) and transcriptomics data [8, 12–14]. These imaging transcriptomic studies often did not directly involve the diagnostic outcome in the analyses. Thus, the identified imaging or transcriptomic markers might not be related or specific to disease outcomes such as AD.

In this study, we hypothesize that the genes whose expression patterns in the brain spatially correlated with the diagnostic pattern of AD-related neuroimaging traits, like beta-amyloid deposition, will also be associated with AD. Thus, we propose an innovative two-stage approach to identify the disease-related genes by correlating spatial gene expression map with regional diagnostic effects on imaging traits in the brain (Fig. 1). We demonstrate the promise of our approach to identifying disease-related gene expression signatures using the brain transcriptome data from the AHBA [15] and the amyloid imaging data from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) cohort [16]. Our approach identifies not only multiple confirmatory findings (i.e., AD genes from the DisGeNET database [17]) but also numerous novel genes associated with various disease stages (i.e., early mild cognitive impairment [EMCI], late mild cognitive impairment [LMCI], and AD); and these novel genes have not been previously linked to AD in DisGeNET.

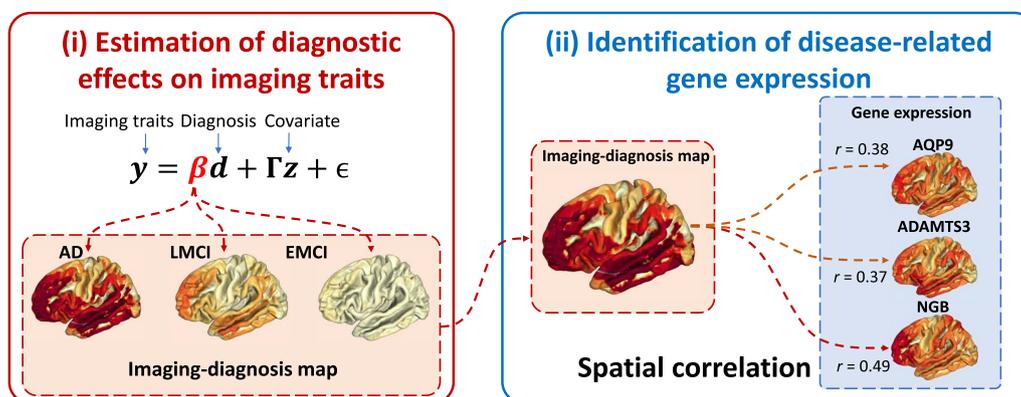


Fig. 1 The overall workflow for identifying novel genes related to Alzheimer’s Disease (AD). The workflow involves: (i) estimating the diagnostic effects on the image traits across the brain using a linear regression model; and (ii) identifying AD-specific genes via spatially correlating imaging-diagnosis map with gene expression map in the brain

Results

In this section, we present the results of our empirical study with the ADNI and AHBA data to demonstrate the promise of our approach. Briefly, in our stage 1 regression analysis, we noted that cognitive normal (CN) vs AD participants yielded the most significant diagnostic effects on the AV45 imaging traits, while CN vs EMCI yielded the least significant ones. In our stage 2 correlation analysis, we identified 12 genes, including 4 AD-related and 8 non-AD-related genes based on the DisGeNET database, whose gene expression maps in the brain were spatially correlated with all the six diagnostic effect maps on AV45 imaging traits. See below for more detailed description of these results.

Mapping diagnostic effects on the AV45 imaging traits across the brain via linear regression

We created an imaging-diagnosis map for each of six diagnostic comparisons through applying a linear regression model to capture the diagnostic effects on AV45 imaging traits across the brain. Figure 2 shows all six brain maps of imaging-diagnosis associations for six different case-control comparisons respectively. The $-\log_{10}(p)$ values for each analysis are color-coded and mapped onto the brain. We observed that CN vs AD yielded the most significant diagnostic effects on the AV45 imaging traits, while CN vs EMCI yielded the least significant ones (Table 1).

Shown in Additional File 1: Table S1 is all the significant ROI-diagnosis associations across the six experiments. In total, 171 out of 180 ROIs were identified to be associated with at least one diagnostic outcome. In addition, 39 ROIs were significantly associated with all six diagnostic outcomes. In detail, 40 ROIs are significant for CN vs EMCI, 145 ROIs are significant for CN vs LMCI, 169 ROIs are significant for CN vs AD, 134 ROIs are significant for CN vs EMCI+LMCI, 167 ROIs are significant for CN vs LMCI+AD, and 162 ROIs are significant for CN vs EMCI+LMCI+AD. These findings are in accordance with our knowledge, where the abnormality of AV45 imaging traits gradually increases over the disease progression from EMCI to LMCI and then to AD.

Table 2 shows the top ten most significant ROIs across the six different diagnostic comparisons. Among these top findings, there are in total 23 ROIs with significant diagnostic effects on the AV45 imaging traits in one or more diagnostic comparisons; and 14 out of 23 ROIs (i.e., L_10pp_ROI, L_10r_ROI, L_10v_ROI, L_31a_ROI, L_31pd_ROI, L_31pv_ROI, L_7m_ROI, L_9a_ROI, L_9p_ROI, L_d23ab_ROI, L_p32_ROI, L_PO_S1_ROI, L_s32_ROI, and L_v23ab_ROI) are subregions of the bilateral dorsomedial parietal lobe, temporal-parietal-occipital junction, and dorsocaudal temporal lobe. Three ROIs (i.e., L_a10p_ROI, L_a47r_ROI, and L_p47r_ROI) are subregions of the frontoparietal cognitive control network.

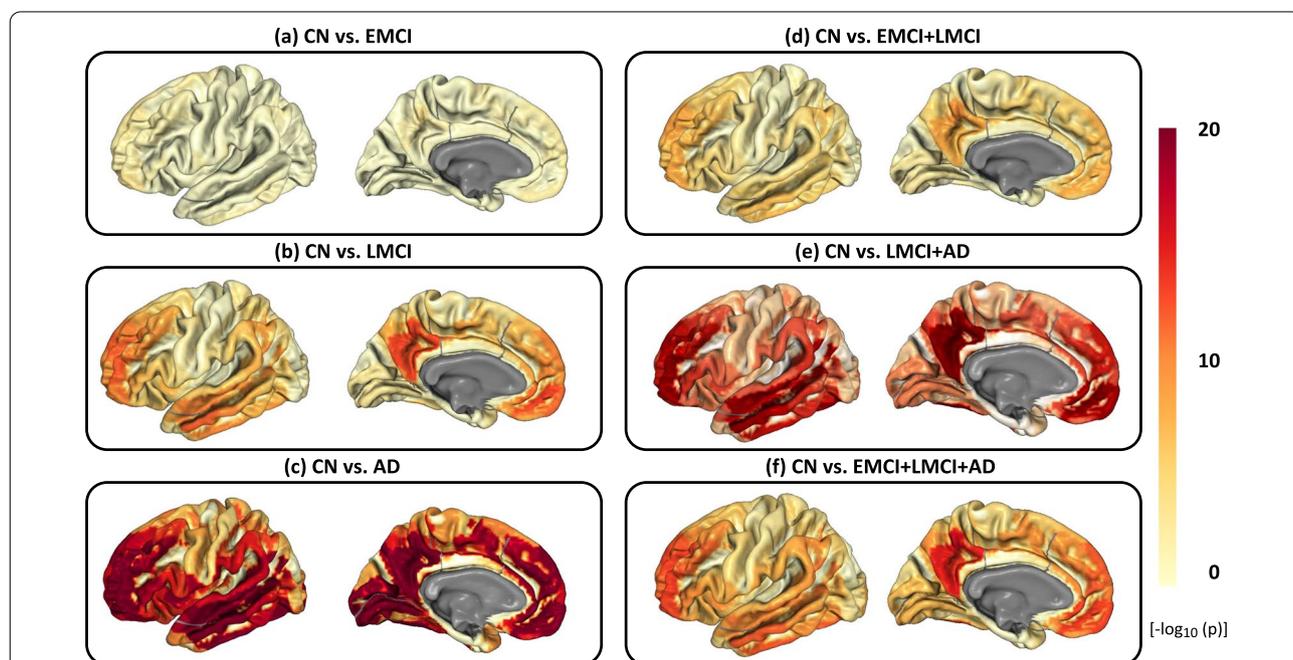


Fig. 2 Brain maps of imaging-diagnosis associations. **a–f** The brain maps for six different case-control comparisons respectively. The $-\log_{10}(p)$ values for each analysis are color-coded and mapped onto the brain

Identifying disease-related genes through brain transcriptome mapping using correlation analysis

We applied Pearson correlation analysis to identify disease-related genes whose expression maps were spatially correlated with the imaging-diagnosis maps. In total, 295 unique genes were identified across the six experiments, and we summarized all these results in Additional File 1: Table S2(a). Specifically, CN vs AD yielded the smallest number of significant correlations with only 40 significant genes, and CN vs LMCI+AD yielded the largest number of significant genes with 214 significant genes.

Out of 295 significant genes, 57 genes have already been linked to AD according to the DisGeNET database [17], and we summarized these 57 gene findings

in Additional File 1: Table S2(b). Figure 3 shows the heat map of these 57 gene findings, where the $-\log_{10}(p)$ values of diagnosis-by-gene correlations are shown between the six diagnostic comparisons on the y axis and all 57 genes on the x-axis. There are four genes (i.e., AQP9, ADAMTS3, HDAC9, and NGB) whose expression maps are significantly correlated with imaging-diagnosis maps for all six diagnostic comparisons. Figure 4 shows the gene expression profiles of the above four genes mapped onto the brain. A visual comparison between six diagnostic effect maps in Fig. 2 and four gene expression maps in Fig. 4 indicates that the spatial patterns of these maps are strongly correlated.

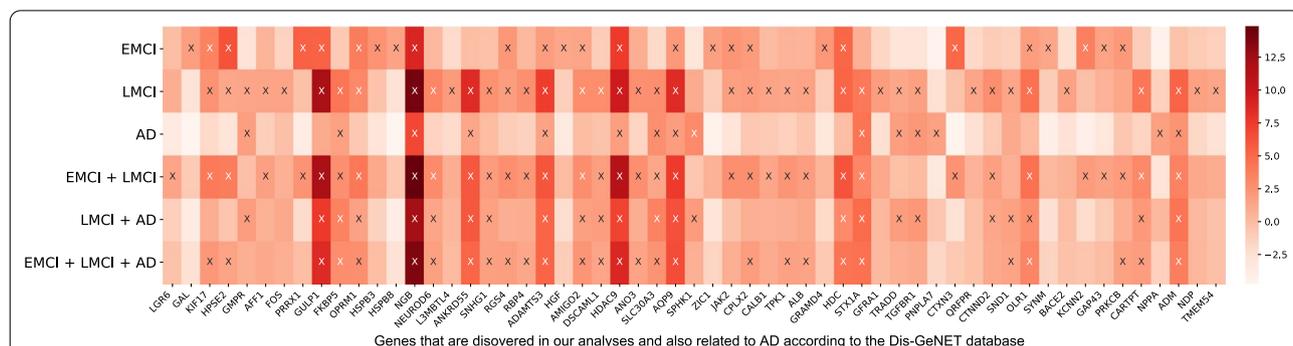


Fig. 3 The heat map of significant gene findings that are related to AD according to Dis-GeNET. The heat map includes all the genes that are discovered in our analyses and also related to AD according to the Dis-GeNET database. The six diagnostic comparisons on the y axis are plotted against all the genes on the x-axis. The $-\log_{10}(p)$ values are color-coded and shown in the heat map. The symbol "X" denotes significant correlation between the imaging-diagnosis map and the gene expression map (corrected $p < 0.05$)

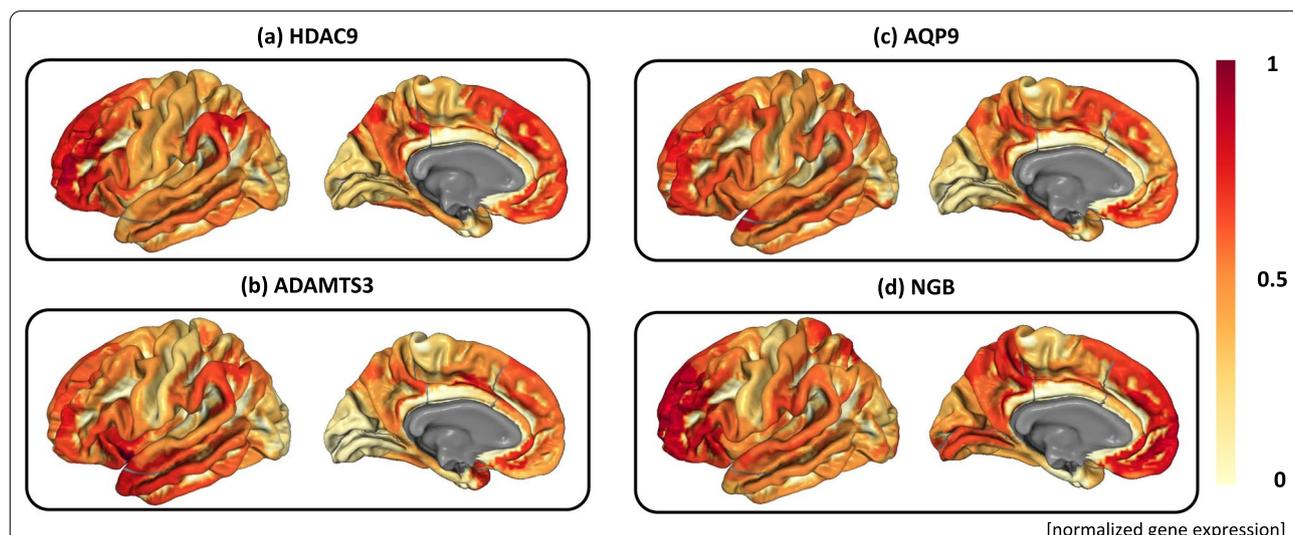


Fig. 4 Gene expression level on the brain for the identified AD-related genes. **a-d** show the brain transcriptome maps for four AD-related genes (i.e., AQP9, ADAMTS3, HDAC9, and NGB) respectively. The normalized gene expression level is color-coded and mapped onto the brain. These maps are significantly correlated with imaging-diagnosis maps for all six case-control comparisons shown in Fig. 2

Figure 5a shows the heat map of four significant gene findings that are not only correlated to all six diagnostic comparisons but also are related to AD according to the DisGeNET database. It is noted that the NGB gene expression pattern has the highest correlations with the imaging-diagnosis maps across the six comparisons (i.e., 0.53 for CN vs EMCI, 0.61 for CN vs LMCI, 0.49 for CN vs AD, 0.62 for CN vs LMCI+AD, 0.59 for CN vs LMCI+AD, and 0.61 for CN vs EMCI+LMCI+AD), and the ADAMTS3 gene expression pattern exhibits the lowest correlations (i.e., 0.38 for CN vs EMCI, 0.50 for CN vs LMCI, 0.37 for CN vs AD, 0.48 for CN vs LMCI+AD, 0.46 for CN vs LMCI+AD, and 0.46 for CN vs EMCI+LMCI+AD).

Out of the 295 genes, there are 238 genes that have not been linked to AD in DisGeNET, denoted as “non-AD-genes”. These 238 gene findings are shown in Additional File 1: Table S2(c). Among these 238 genes, there are 8 genes (i.e., ASB2, NPY1R, GLRA3, COL23A1, SPRN, CPNE8, TSPAN33, and KCNA3) whose expression maps are significantly correlated with imaging-diagnosis maps for all six diagnostic comparisons. These findings suggest potential connections between these genes and AD through underlying transcriptomic and/or amyloid mechanisms in the brain, which is an interesting research topic warranting further investigation. Figure 5b shows the heat map of these 8 significant gene findings. The SPRN gene has the highest correlation coefficients (i.e., 0.47 for CN vs EMCI, 0.52 for CN vs LMCI, 0.42 for CN vs AD, 0.53 for CN vs LMCI+AD, 0.50 for CN vs LMCI+AD, and 0.52 for CN vs EMCI+LMCI+AD), and the NPY1R gene has the lowest correlation coefficients (i.e., 0.37 for CN vs EMCI, 0.40 for CN vs LMCI, 0.36

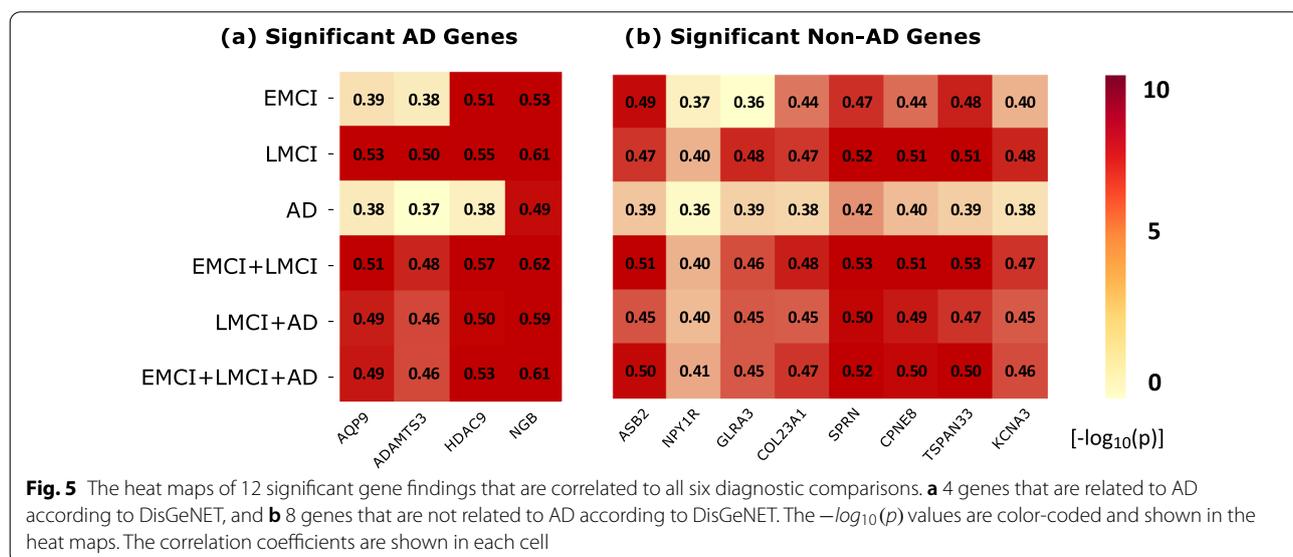
for CN vs AD, 0.40 for CN vs LMCI+AD, 0.40 for CN vs LMCI+AD, and 0.41 for CN vs EMCI+LMCI+AD).

Among 12 gene findings shown in Fig. 5a, b, we observed that CN vs AD yielded the lowest correlation coefficients for all 12 genes, in comparison with all the other five diagnostic paradigms. This pattern appears intriguing, and warrants replication in independent cohorts as well as further investigation for possible mechanistic understanding.

Discussion

In this work, we have proposed a two-stage approach to link transcriptome data to diagnosis outcome via brain mapping and have applied it to the AHBA and ADNI data to demonstrate its promise. Many studies have performed correlation or regression analysis to investigate the relationship among genetics, transcriptomics and neurodegenerative disorders [7]. To the best of our knowledge, this work is the first study to link diagnostic outcomes and transcriptomic data via human brain mapping. Across the six experiments, we confirmed the following four genes that are linked to AD in the DisGeNET database: AQP9, ADAMTS3, HDAC9, and NGB, as shown in Fig. 5a. In addition, we identified the following eight novel genes not yet linked to AD according to DisGeNET: ASB2, NPY1R, GLRA3, COL23A1, SPRN, CPNE8, TSPAN33, and KCNA3, as shown in Fig. 5b.

We observed that four significant AD genes were correlated with all six case-control comparisons—AQP9, ADAMTS3, HDAC9, and NGB. AQP9 was reportedly expressed in the neurons of the substantia nigra, tanyocytes, and some astrocytes [18]. One mouse study demonstrated that the reduction of ADAMTS3 contributed



to the inhibition of amyloid β decomposition [19]. One study demonstrated that HDAC9 was reported to be part of the pathway that mediates synaptic function and amyloid precursor protein processing in AD [20]. The expression of NGB was also reported to be down-regulated with increasing age, down-regulated in women (consistent with their increased risk), and up-regulated in the temporal lobe of AD patients (consistent with a response to the disease process) [21]. All these existing findings collectively support our discoveries and demonstrate the effectiveness of our method in identifying AD-related genes supported by the transcriptomic evidence in the brain.

Additionally, we observed that eight non-AD genes significantly correlated with the image-diagnosis map for all six case-control comparisons—ASB2, NPY1R, GLRA3, COL23A1, SPRN, CPNE8, TSPAN33, and KCNA3. We used the statistical over-representation test of PANTHER [22] to detect statistical over-representation of the eight significant non-AD genes compared to the human genome reference gene list. In this process, the enriched biological process of *Homo sapiens* species via gene ontology (GO) [23, 24] was identified. The results showed that 28 GO biological processes passed the nominal threshold of $p < 0.05$. Among all these significant GO biological processes, “neuropeptide signaling pathway” was enriched with the smallest raw p -value 1.01×10^{-3} and with fold enrichment 41.99. Existing findings, including apeline, Neuropeptide Y, and dynorphin A, which were important neuropeptides, were closely related to AD [25–27]. Moreover, evidence showed that in AD patients and AD animal models, numerous neuropeptide-containing neurons were pathologically altered in their brain areas [28]. The levels of various neuropeptides had also been found altered in both Cerebrospinal Fluid, and blood of AD patients [28]. All these existing findings demonstrate an important role of “neuropeptide signaling pathway” played in AD, indicating that the eight significant DisGeNET non-AD genes might have a connection to AD. It warrants further investigation to evaluate the potential of these genes as AD related genes.

Many AD studies demonstrated that the accumulation of misfolded amyloid β and tau protein causes degeneration and loss of neuronal function in the brain [29, 30]. Our analyses yielded promising gene findings through mapping gene expression pattern and AD-related amyloidosis pattern in the brain. Our study not only confirmed known AD genes but also suggested novel gene targets with potential to be linked to AD through underlying transcriptomic and amyloid mechanisms in the brain. These discoveries provide valuable information to guide further neurobiological and molecular studies to reveal the biological pathway from gene expression to amyloid

accumulation and to the degeneration and loss of neuronal function in the brain.

Conclusion

We have proposed a two-stage approach to identify genes whose expression levels are related to diagnosis phenotype via brain transcriptome mapping. Specifically, we first mapped the effects of a diagnosis phenotype onto imaging traits across the brain using a linear regression model. Then, the gene-diagnosis association was assessed by spatially correlating the brain transcriptome map with the diagnostic effect map on the brain-wide imaging traits. We applied our approach to analyze the brain transcriptome data from the AHBA and the amyloid imaging data from the ADNI cohort to demonstrate the promise of our approach. Our approach detected four genes with AD link shown in DisGeNET and eight genes that have not yet been linked to AD in DisGeNET. Our proposed novel disease-related brain transcriptomic mapping method was designed to spatially associate gene expression profiles with regional diagnostic effects on a brain trait to reveal disease-related genes. Our empirical study on the AHBA and ADNI data shows the promise of the approach. The resulting AD gene discoveries provide valuable information to better understand biological pathways from transcriptomic signatures to intermediate brain traits and phenotypic disease outcomes.

Methods

Data description

Data used in the preparation of this article were obtained from the ADNI database (adni.loni.usc.edu) [16]. The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. 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). For up-to-date information, see www.adni-info.org.

In this study, we included a total of 971 participants (i.e., 255 Cognitive Normal [CN], 296 EMCI, 218 LMCI, and 202 AD subjects) who had complete [^{18}F]florbetapir (AV45) PET data (measuring amyloid burden), and diagnostic and clinical assessments. The detailed demographic information is presented in Table 1. For AV45 PET scans, the data was registered to the Montreal Neurological Institute space, and the standard uptake value ratio was computed by intensity normalization using the cerebellar *curs* reference region. Region-level AV45 measures were extracted based on the human

Table 1 Demographic information

	CN	EMCI	LMCI	AD	Total
Number of subject	255	296	218	202	971
Age	76.35 ± 6.54	71.78 ± 7.28	74.71 ± 8.39	75.85 ± 7.67	74.48 ± 7.67
Sex (male/female)	132/123	167/129	129/89	123/79	551/420
Education (year)	16.37 ± 2.64	12.12 ± 2.64	16.12 ± 2.94	15.83 ± 2.81	16.13 ± 2.75

Table 2 Top 10 significant regions associated with diagnostic outcome

ROI	Network	CN versus EMCI	CN versus LMCI	CN versus AD	CN versus EMCI + LMCI	CN versus LMCI + AD	CN versus EMCI + LMCI + AD
L_10pp_ROI	Posterior-multimodal	3.21			8.08		
L_10r_ROI	Posterior-multimodal		12.86	28.06		20.61	
L_10v_ROI	Posterior-multimodal	2.15					
L_31a_ROI	Posterior-multimodal	2.05	12.63		8.48	20.60	14.36
L_31pd_ROI	Posterior-multimodal		11.98	29.56		22.35	13.89
L_31pv_ROI	Posterior-multimodal		12.31	30.03		22.48	13.48
L_6ma_ROI	Cingulo-opercular	2.03					
L_7m_ROI	Posterior-multimodal		12.76	30.00	7.96	22.92	14.51
L_9-46d_ROI	Cingulo-opercular				7.79		13.53
L_9a_ROI	Posterior-multimodal	2.54			8.48		
L_9p_ROI	Posterior-multimodal	2.42			7.81		
L_a10p_ROI	Frontoparietal	2.03					
L_a47r_ROI	Frontoparietal	3.00					
L_d23ab_ROI	Posterior-multimodal		13.96	32.99	8.76	24.26	15.72
L_FFC_ROI	Visual2			30.42		21.20	
L_p32_ROI	Posterior-multimodal		12.64				
L_p47r_ROI	Frontoparietal	2.48			7.99		13.22
L_PCV_ROI	Dorsal-attention	3.15	12.91		9.55	21.68	15.79
L_POS1_ROI	Posterior-multimodal			37.73		22.03	
L_s32_ROI	Posterior-multimodal		12.53		7.89		13.33
L_TE2p_ROI	Language			29.02			
L_v23ab_ROI	Posterior-multimodal		12.75	37.54		24.64	14.89
L_VVC_ROI	Visual2			33.08			

The significance level is reported in the format of $-\log_{10}(p)$ -value. The network is annotated to the each ROI according to the Cole-Anticevic Brain-side Network [32]

connectome project [31] multi-modal parcellation (HCP-MMP) atlas, consisting of 360 regions of interests (ROIs).

The brain-wide transcriptome (gene expression) data was downloaded from the Allen Human Brain Institute [15]. The AHBA offered the gene expression levels of more than 20,000 genes from 3,702 distinct tissues, which were sampled across six different donors to cover the entire brain. Of these six donors, only two donors had tissue samples from both hemispheres, and the remaining four donors had the samples only from the left hemisphere. In this work, we preprocessed the AHBA data using the pipeline proposed by Aurina *et al.* and mapped those onto the HCPMMP atlas [8]. After that, we focused

our analyses on the data from the left hemisphere (including 180 ROIs), since it is more densely sampled than the right one.

Two-stage method for identifying novel genes related to Alzheimer's disease

Here we propose a two-stage method for identifying novel genes related to Alzheimer's disease, as summarized in Fig. 1. In the first stage, we perform linear regression to estimate the diagnostic effects on the image traits across the brain. In the second stage, we identify novel genes whose gene expression levels in the brain

are spatially correlated with the diagnostic effects on the imaging traits across the brain.

Stage 1. Regression analysis to estimate diagnostic effect on the AV45 imaging traits in the brain: imaging-diagnosis map

Utilizing the AV45 imaging traits and diagnostic phenotype, we performed linear regression analysis to estimate the diagnostic effect on imaging traits. In this analysis, we used the ROI-based imaging trait as the response, the diagnostic phenotype (i.e., control vs case) as the predictor, and age, sex, and education level as covariates to remove confounding effects. Since AD is a progressive neurodegenerative disorder, it is important to perform the analysis for different disease stages, where the disease progresses from EMCI to LMCI and then to AD (i.e., EMCI < LMCI < AD). With this observation, we examined the following six diagnostic comparisons: 1) CN vs EMCI, 2) CN vs LMCI, 3) CN vs AD, 4) CN vs EMCI+LMCI, 5) CN vs LMCI+AD, and 6) CN vs EMCI+LMCI+AD. Of note, CN vs EMCI+AD comparison was excluded due to the stage discontinuity between EMCI and AD (i.e., LMCI is in-between). We applied our approach to each of the above six cases to estimate the corresponding diagnostic effect on the brain traits. As a result, we generated the significance map (imaging-diagnosis map) to assess the diagnostic effects in the brain. The significance was reported in format of $-\log_{10}(p\text{-value})$. We then excluded all the ROIs in the right hemisphere, to solely focus on the left hemisphere for the reason we stated earlier.

Stage 2. Correlation analysis between gene expression map and imaging-diagnosis map

In this section, we identify disease-related genes whose expression levels are spatially correlated to imaging-diagnosis map through computing their Pearson's correlation coefficients. We performed this analysis 60,162 times (i.e., 6 imaging-diagnosis maps \times 10,027 genes expression maps from the AHBA database) and the Bonferroni method was applied for multiple comparison correction. Due to the preprocessing step for the AHBA data, we only utilized the left hemisphere (i.e., 180 ROIs out of the total 360 ROIs) for both imaging-diagnosis map and the AHBA transcriptome map. In addition, we further analyzed the identified genes to examine their associations with AD based on the DisGeNET database [17]. The DisGeNET's comprehensive database for disease association studies was utilized to assess whether or not the significant genes that we have identified have been associated with AD in prior studies. In DisGeNET, there are 3397 genes related to "Alzheimer's Disease" (UMLS CUI: C0002395). Out of

these 3397 genes, we only looked at a subset of 1877 genes since the pre-processing step for the AHBA data removed genes with insignificant expression.

Abbreviations

AD: Alzheimer's disease; ADNI: Alzheimer's disease neuroimaging initiative; ROI: Region of interest; AHBA: Allen Human Brain Atlas human brain microarray data; AV45: [^{18}F]florbetapir; EMCI: Early mild cognitive impairment; LMCI: Late mild cognitive impairment; CN: Cognitively normal.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12920-022-01260-6>.

Additional file 1. Table S1: Significance level of brain regions using imaging-diagnosis analysis. **Table S2(a):** Significance level of identified genes significantly associated with diagnostic outcome using correlation analysis. **Table S2(b):** Significance level of identified genes related with AD based on DisGeNET. **Table S2(c):** Significance level of identified genes unrelated with AD based on DisGeNET.

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.; Cogstate; 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 Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

Alzheimer's Disease Neuroimaging Initiative—Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.ucla.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. 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

Author contributions

MS, JYB, JX, QL and LS: conceptualization and study design. MS: data preprocessing. MS and JYB: method implementation and data analysis. MS, JYB, JX and LS: manuscript drafting. All the authors reviewed, commented, revised and approved the manuscript.

Funding

This work and publication costs are supported by the National Institutes of Health [R01 LM013463, U01 AG068057, RF1 AG063481, R01 AG058854] and the National Research Foundation of Korea [NRF-2020R1A6A3A03038525]. The funders were not involved in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Availability of data and materials

The datasets used and analyzed during the study are available in the ADNI LONI repository and Allen brain atlas data portal, <https://adni.loni.usc.edu/> and <https://portal.brain-map.org/>, respectively.

Declarations**Ethics approval and consent to participate**

This research is conducted under the regulation of Institutional Review Boards (IRB) and the research subject informed consent process at University of Pennsylvania, USA. Study subjects gave written informed consent at the time of enrollment for data collection and completed questionnaires approved by each participating site's IRB. The authors state that they have obtained approval from the Alzheimer's Disease Neuroimaging Initiative (ADNI) Data Sharing and Publications Committee for use of the data.

Consent for publication

Not applicable. Secondary analysis of publicly available data

Competing interests

The authors declare that they have no competing interests.

Author details

¹School of Arts and Sciences, University of Pennsylvania, Philadelphia, USA. ²Department of Artificial intelligence, Catholic University of Korea, Bucheon, Republic of Korea. ³Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, USA.

Received: 2 May 2022 Accepted: 4 May 2022

Published online: 19 May 2022

References

- Gatz M, Reynolds CA, Fratiglioni L, Johansson B, Mortimer JA, Berg S, Fiske A, Pedersen NL. Role of genes and environments for explaining Alzheimer disease. *Arch Gen Psychiatry*. 2006;63(2):168–74.
- Hollingworth P, Harold D, Sims R, Gerrish A, Lambert J-C, Carrasquillo MM, Abraham R, Hamshere ML, Pahwa JS, Moskvina V, et al. Common variants at *abca7*, *ms4a6a/ms4a4e*, *epha1*, *cd33* and *cd2ap* are associated with Alzheimer's disease. *Nat Genet*. 2011;43(5):429–35.
- Lambert J-C, Heath S, Even G, Campion D, Sleegers K, Hiltunen M, Combarros O, Zelenika D, Bullido MJ, Tavernier B, et al. Genome-wide association study identifies variants at *clu* and *cr1* associated with Alzheimer's disease. *Nat Genet*. 2009;41(10):1094–9.
- Hu X, Pickering E, Liu YC, Hall S, Fournier H, Katz E, Dechairo B, John S, Van Eerdewegh P, Soares H, et al. Meta-analysis for genome-wide association study identifies multiple variants at the *bin1* locus associated with late-onset Alzheimer's disease. *PLoS ONE*. 2011;6(2):16616.
- Khanha S, Domingo-Fernández D, Iyappan A, Emon MA, Hofmann-Apitius M, Fröhlich H. Using multi-scale genetic, neuroimaging and clinical data for predicting Alzheimer's disease and reconstruction of relevant biological mechanisms. *Sci Rep*. 2018;8(1):1–13.
- Mroczek M, Desouky A, Sirry W. Imaging transcriptomics in neurodegenerative diseases. *J Neuroimaging*. 2021;31(2):244–50.
- Shen L, Thompson PM. Brain imaging genomics: integrated analysis and machine learning. *Proc IEEE Inst Electr Electron Eng*. 2020;108(1):125–62.
- Arnatkeviciute A, Fulcher BD, Fornito A. A practical guide to linking brain-wide gene expression and neuroimaging data. *Neuroimage*. 2019;189:353–67.
- Selvaggi P, Rizzo G, Mehta MA, Turkheimer FE, Veronese M. Integration of human whole-brain transcriptome and neuroimaging data: practical considerations of current available methods. *J Neurosci Methods*. 2021;355:109128.
- Nho K, Nudelman K, Allen M, Hodges A, Kim S, Risacher SL, Apostolova LG, Lin K, Lunnon K, Wang X, et al. Genome-wide transcriptome analysis identifies novel dysregulated genes implicated in Alzheimer's pathology. *Alzheimer's Dement*. 2020;16(9):1213–23.
- Berchtold NC, Sabbagh MN, Beach TG, Kim RC, Cribbs DH, Cotman CW. Brain gene expression patterns differentiate mild cognitive impairment from normal aged and Alzheimer's disease. *Neurobiol Aging*. 2014;35(9):1961–72.
- Freeze B, Acosta D, Pandya S, Zhao Y, Raj A. Regional expression of genes mediating trans-synaptic alpha-synuclein transfer predicts regional atrophy in Parkinson disease. *NeuroImage Clin*. 2018;18:456–66.
- Yu L, Dawe RJ, Boyle PA, Gaiteri C, Yang J, Buchman AS, Schneider JA, Arfanakis K, De Jager PL, Bennett DA. Association between brain gene expression, dna methylation, and alteration of ex vivo magnetic resonance imaging transverse relaxation in late-life cognitive decline. *JAMA Neurol*. 2017;74(12):1473–80.
- Acosta D, Powell F, Zhao Y, Raj A. Regional vulnerability in Alzheimer's disease: The role of cell-autonomous and transneuronal processes. *Alzheimer's Dement*. 2018;14(6):797–810.
- Hawrylycz MJ, Lein ES, Guillozet-Bongaerts AL, Shen EH, Ng L, Miller JA, Van De Lagemaat LN, Smith KA, Ebbert A, Riley ZL, et al. An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature*. 2012;489(7416):391–9.
- Weiner MW, Veitch DP, Aisen PS, Beckett LA, Cairns NJ, Green RC, Harvey D, Jack CR, Jagust W, Liu E, Morris JC, Petersen RC, Saykin AJ, Schmidt ME, Shaw L, Shen L, Siuciak JA, Soares H, Toga AW, Trojanowski JQ. Alzheimer's disease neuroimaging, I. The Alzheimer's disease neuroimaging initiative: a review of papers published since its inception. *Alzheimer's Dement*. 2013;9(5):111–94.
- Piñero J, Bravo À, Queralt-Rosinach N, Gutiérrez-Sacristán A, Deu-Pons J, Centeno E, García-García J, Sanz F, Furlong LI. Disgenet: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res*. 2016;66:943.
- Papadopoulos M, Verkman A. Aquaporin water channels in the nervous system. *Nat Rev Neurosci*. 2013;66:14.
- Yamagake Y, Tsujii H, Kohno T, Ogino H, Saito T, Saido TC, Hattori M. Reducing *adams-3* inhibits amyloid β deposition in app knock-in mouse. *Biol Pharm Bull*. 2019;42(3):354–6.
- Lu Y, Tan L, Wang X. Circular *hdac9/microrna-138/sirtuin-1* pathway mediates synaptic and amyloid precursor protein processing deficits in Alzheimer's disease. *Neurosci Bull*. 2019;35:66.
- Szymanski M, Wang R, Fallin MD, Bassett SS, Avramopoulos D. Neuroglobin and Alzheimer's dementia: genetic association and gene expression changes. *Neurobiol Aging*. 2010;31(11):1835–42.
- Mi H, Muruganujan A, Ebert D, Huang X, Thomas PD. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res*. 2018;47(D1):419–26.
- Ashburner M, Ball CA, Blake J, Botstein D, Butler H, Cherry J. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*. 2000;25:25–9.
- Consortium TGO. The Gene Ontology resource: enriching a GOLD mine. *Nucleic Acids Res*. 2020;49(D1):325–34.
- Masoumi J, Abbasloui M, Parvan R, Mohammadnejad D, Pavon-Djavid G, Barzegari A, Abdolalizadeh J. Apelin, a promising target for Alzheimer disease prevention and treatment. *Neuropeptides*. 2018;70:76–86.
- Spencer B, Potkar R, Metcalf J, Thrin I, Adame A, Rockenstein E, Masliah E. Systemic central nervous system (cns)-targeted delivery of neuropeptide γ (*np γ*) reduces neurodegeneration and increases neural precursor cell proliferation in a mouse model of Alzheimer disease*. *J Biol Chem*. 2016;291(4):1905–20.
- Yakovleva T, Marinova Z, Kuzmin A, Seidah NG, Haroutunian V, Terenius L, Bakalkin G. Dysregulation of dynorphins in Alzheimer disease. *Neurobiol Aging*. 2007;28(11):1700–8.
- Petrella C, Certo M, Barbato C, Gabanella F, Ralli M, Greco A, Possenti R, Severini C. Neuropeptides in Alzheimer's disease: an update. *Curr Alzheimer Res*. 2019;16:66.
- Pasieka A, Panek D, Szalaj N, Espargaró A, Wieckowska A, Malawska B, Sabaté R, Bajda M. Dual inhibitors of amyloid- β and tau aggregation with amyloid- β disaggregating properties: extended in cellulo, in silico, and kinetic studies of multifunctional anti-Alzheimer's agents. *ACS Chem Neurosci*. 2021;6:66.
- Heller GT, Aprile FA, Michaels TC, Limbocker R, Perni M, Ruggeri FS, Manini B, Löhr T, Bonomi M, Camilloni C, et al. Small-molecule sequestration of amyloid- β as a drug discovery strategy for Alzheimer's disease. *Sci Adv*. 2020;6(45):5924.

31. Van Essen DC, Smith SM, Barch DM, Behrens TE, Yacoub E, Ugurbil K, Consortium W-MH, et al. The Wu-Minn human connectome project: an overview. *Neuroimage*. 2013;80:62–79.
32. Ji JL, Spronk M, Kulkarni K, Repovš G, Anticevic A, Cole MW. Mapping the human brain's cortical–subcortical functional network organization. *Neuroimage*. 2019;185:35–57.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

