Polygenic mediation analysis of Alzheimer's disease implicated intermediate amyloid imaging phenotypes

Yingxuan Eng, HS[a,1], Xiaohui Yao, PhD[a,1], Kefei Liu, PhD[a], Shannon L. Risacher, PhD[b], Andrew J. Saykin, PsyD[b], Qi Long, PhD[a], Yize Zhao, PhD[c], Li Shen, PhD[a,2], for the ADNI[3]
[a]University of Pennsylvania, Philadelphia, PA 19104, USA; [b]Indiana University, Indianapolis, IN 46202, USA; [c]Yale University, New Haven, CT 06511, USA

Abstract

Mediation models have been employed in the study of brain disorders to detect the underlying mechanisms between genetic variants and diagnostic outcomes implicitly mediated by intermediate imaging biomarkers. However, the statistical power is influenced by the modest effects of individual genetic variants on both diagnostic and imaging phenotypes and the limited sample sizes of imaging genetic cohorts. In this study, we propose a polygenic mediation analysis that comprises a polygenic risk score (PRS) to aggregate genetic effects of a set of candidate variants and then explore the implicit effect of imaging phenotypes between the PRS and disease status. We applied our proposed method to an amyloid imaging genetic study of Alzheimer's disease (AD), identified multiple imaging mediators linking PRS with AD, and further demonstrated the promise of the PRS on mediator detection over individual variants alone.

Introduction

Genome-wide association study (GWAS) of complex brain disorders have discovered a few genetic risk variants for diseases. For example, various case-control GWAS have been performed in Alzheimer's disease (AD) and identified multiple AD susceptible loci[1–3]. However, given the increasingly divergent and complex path from genes to disease, it is a challenge to directly link genetic risk variants with brain disorders[4]. Imaging genetics, a rapidly growing research field, has been developed to investigate the genetic effects on brain structures and functions, which represent a more proximate biological link to genes and serve as obligatory intermediate of cognitive and behavioral outcomes. For example, the amyloid accumulation in brain measured by AV45-PET imaging, which is also referred as "A" in the "A/T/N" classifcation scheme of AD[5], is one of the most promising biomarkers for diagnosis and classification of AD. Imaging genetics[6] has identified a few individual risk loci for brain imaging quantitative traits (iQTs)[2], while the gap between iQT associated variants with brain disorders still needs to be bridged - that is, incorporating brain imaging information into the genetic analysis of brain disorders can help provide new insights into the phenotypic characteristics and the implicit molecular mechanisms of disordered cognitive and behavioral outcomes.

Recently, mediation analysis, a statistic model aiming to identify the underlying mechanism of an observed relationship between an independent variable and a dependent variable through a third hypothetical variable (i.e., mediator variable), has been employed in genetic analysis of brain disorders to detect mechanisms between genetic variants and disordered outcomes implicitly mediated by brain imaging phenotypes. For example, Bi et al.[7] performed a genome-wide mediation analysis of cognitive traits using neuroimaging measures as intermediates, and successfully detected several genetic variants for their implicit effects on cognitive behaviors mediated through regional brain structures.

Mediation analysis requires associations of the independent variable with the dependent variable and with the mediator to be both significant. This makes applying it in brain disorder studies a challenge due to the modest effect of an individual genetic variant on both diagnostic and imaging phenotypes[8], as well as the limited size of the study sample with all genetic, imaging and diagnostic data available. Moreover, complex diseases are typically influenced by a collective effect from multiple variants, instead of by a single variant alone. Hence, multivariate association approaches involving multiple single nucleotide polymorphisms (SNPs) have been developed to consider the joint effect of a set of SNPs, to increase the statistical power and biological interpretation.

[1]These authors contributed equally to this work.
[2]Correspondence: Li.Shen@pennmedicine.upenn.edu
[3]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 data analysis or writing of this report. A complete listing of ADNI investigators can be found at: `https://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf`

Table 1: Participant characteristics in AV45-PET mediation analysis.

Diagnosis	HC	EMCI	LMCI	AD	p-value
Number	204	246	169	140	-
Gender(M/F)	100/104	128/118	90/79	76/64	7.72E-01
Age(mean±std)	75.70±6.46	71.33±7.32	73.78±8.72	75.10±8.01	**4.20E-09**
Education(mean±std)	16.32±2.72	16.04±2.63	16.18±2.87	15.64±2.69	1.39E-01
APOE ε4 present	27.59%	42.04%	49.11%	67.14%	**8.13E-12**

P-values were assessed to examine whether the differences among diagnosis groups are significant, and were computed using one-way ANOVA (except for gender using χ^2 test). The p values < 0.05 are shown in bold. HC = Healthy Control; EMCI = Early Mild Cognitive Complaint; LMCI = Late Mild Cognitive Complaint; AD = Alzheimer's Disease

Polygenic risk score (PRS)[9] is one of the popularly used multivariate association approaches, and has been widely applied to many complex diseases, for example, Alzheimer's disease (AD)[10], to understand the underlying genetic architecture. PRS aggregates the genetic effects of a set of trait-related SNPs which may not be individually significant while may cumulatively contribute to the phenotypic variance. The PRS is typically calculated using the effect sizes of SNPs from a large genetic association analysis of diagnostic traits to improve the power of detecting genetic factors. For example, PRS studies of AD have been conducted by including significant SNPs from large AD meta-analysis, and have successfully demonstrated the contribution of their collective effect on AD and its biomarkers[10].

In this study, we propose to explore the intermediate role of imaging QTs in the study of AD. Given the modest size of available samples having both imaging and genetic data, we perform a polygenic mediation analysis of AD to detect brain imaging mediators which intermediately link the pathological path from gene to disease. Specifically, we construct a PRS using a set of candidate disease SNPs with corresponding genetic effects obtained from a large meta-analysis of AD, and apply the mediation model to detect the indirect effect of the PRS on disease via imaging phenotypes. To show the effectiveness of our implemented polygenic mediation analysis, we compare the performance of PRS with the well-known AD SNPs in the *APOE* gene on imaging mediator identification, as well as on the detection of imaging mediators for their abilities of early stage prediction.

Materials and Methods

To demonstrate the power of polygenic mediation analysis for identifying imaging modulators, we apply it to the amyloid imaging genetic analysis in the study of AD. "Amyloid cascade hypothesis" has been considered the leading pathogenesis of AD for decades where brain amyloid deposition is thought to happen over years before the early symptom of AD[11,12], and can be measured by brain imaging methods.

This study was approved by institutional review boards of all participating institutions and written informed consent was obtained from all participants or authorized representatives.

Alzheimer's Disease Neuroimaging Initiative

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu)[13]. 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 AD. For up-to-date information, see `www.adni-info.org`.

AV45-PET data acquisition and processing

Preprocessed [^{18}F]Florbetapir (or ^{18}F-AV-45) PET scans (i.e., amyloid imaging data) were downloaded from the ADNI website (`adni.loni.usc.edu`), then aligned to the corresponding MRI scans and normalized to the MNI space as $2 \times 2 \times 2$ mm voxels. Amyloid measures from 116 regions of interest (ROIs) were further extracted based on the MarsBaR AAL atlas[14]. We excluded the cerebellar ROIs due to their lack of flubetapir tracer activity and

Table 2: Details of AD candidate SNPs used to construct two polygenic risk scores PRS+ and PRS-: (1) PRS+ was computed from all 23 SNPs, and (2) PRS- was computed from 22 SNPs without the *APOE* SNP rs41289512.

Closest gene	CHR	SNP	BP	A1	A2	Meta-GWAS beta	Meta-GWAS p
ADAMTS4	1	rs4575098	161155392	A	G	0.016	2.05E-10
CR1	1	rs2093760	207786828	A	G	0.024	1.10E-18
BIN1	2	rs4663105	127891427	C	A	0.030	3.38E-44
INPPD5	2	rs10933431	233981912	G	C	-0.015	8.92E-10
CLNK	4	rs6448453	11026028	A	G	0.014	1.93E-09
HS3ST1	4	rs7657553	11723235	A	G	0.005	5.09E-02
CD2AP	6	rs9381563	47432637	C	T	0.014	2.52E-10
ZCWPW1	7	rs1859788	99971834	A	G	-0.018	2.22E-15
EPHA1	7	rs7810606	143108158	T	C	-0.014	3.59E-11
CLU/PTK2B	8	rs4236673	27464929	A	G	-0.020	2.61E-19
ECHDC3	10	rs11257238	11717397	C	T	0.013	1.26E-08
MS4A6A	11	rs2081545	59958380	A	C	-0.017	1.55E-15
PICALM	11	rs867611	85776544	G	A	-0.020	2.19E-18
SLC24A4	14	rs12590654	92938855	A	G	-0.014	1.65E-10
ADAM10	15	rs442495	59022615	C	T	-0.013	1.31E-09
KAT8	16	rs59735493	31133100	A	G	-0.013	3.98E-08
SCIMP	17	rs113260531	5138980	A	G	0.019	9.16E-10
ABI3	17	rs28394864	47450775	A	G	0.012	1.87E-08
BZRAP1-AS1	17	rs2632516	56409089	C	G	-0.010	9.66E-07
ABCA7	19	rs111278892	1039323	G	C	0.019	7.93E-11
APOE	19	rs41289512	45351516	G	C	0.200	5.79E-276
CD33	19	rs3865444	51727962	A	C	-0.013	6.34E-09
CASS4	20	rs6014724	54998544	G	A	-0.022	6.56E-10

finally included 90 amyloid imaging measures in our analysis. 759 non-Hispanic Caucasian participants (Table 1) with complete baseline ROI-level AV45 measurements were studied.

Genotyping data

Genotyping data were also obtained from ADNI, and were quality controlled (QCed) as described previously[15]. Briefly, genotyping was performed on all ADNI participants following manufacturer's protocol using blood genomic DNA samples and Illumina GWAS arrays (610-Quad; OmniExpress, or HumanOmni2.5-4v1)[16]. QC was performed in PLINK v1.90[17] using the following criteria: 1) call rate per marker $\geq$ 95%, 2) minor allele frequency $\geq$ 5%, 3) Hardy Weinberg Equilibrium test p $\geq$ 1.0E-6, and 4) call rate per participant $\geq$ 95%. Significant relatedness pairs with PI_HAT $>$ 0.45 were identified and thereafter one individual from each pair was randomly excluded. Haplotype patterns from the 1,000 Genomes Project reference panel were then applied to impute the SNPs that were not directly genotyped from arrays. In total, 5,574,300 SNPs were obtained for 759 participants involved in this work.

Polygenic risk scores

A polygenic risk score (PRS) is typically calculated based on a set of trait-related SNPs as the sum of their genotype values (i.e., numbers of minor alleles) weighted by their corresponding effect sizes on the diagnostic trait. In this work, we propose to comprise PRSs using AD candidate SNPs. A most recent large scale meta analytical GWAS (meta-GWAS) of clinically diagnosed AD and AD-by-proxy with totally $455,258$ samples (Phase 3, $71,880$ cases and $383,378$ controls)[3] was conducted and identified a list of 32 AD SNPs in addition to *APOE* rs429358, which is the best known AD genetic risk factor. Among these 32 SNPs, there were 23 available in our genotyping data. These 23 AD candidate SNPs were employed in this study for PRS calculation. The summary statistics of all these AD candidate SNPs were downloaded from `https://ctg.cncr.nl/software/summary_statistics`. Table 2 shows the detailed information of these 23 SNPs, including their genetic effects on AD.

Given the 23 candidate SNPs, we extracted their genotype values from the genotyping data of 759 ADNI participants and obtained their effect sizes on AD from the large meta-GWAS of AD. Below we describe how a PRS is computed based on a set of AD candidate SNPs. We use $\mathbf{G} = [g_{i,j}]$ to denote the genotype data from ADNI, where $g_{i,j} \in (0, 1, 2)$ is the number of minor alleles of SNP j in subject i. For subject i, given a set of AD candidate SNPs $S = \{s_1, \ldots, s_n\}$, the PRS is calculated as follows:

$$\text{PRS}_i = \sum_{j=1}^{n} \beta_j \cdot \text{g}_{i,j}, \tag{1}$$

where β_j is the beta coefficient of SNP j obtained from the meta-GWAS of AD. We then standardized PRS across subjects to have mean of zero and standard deviation of one.

To discover novel signals contributed by SNPs in addition to the well-known *APOE* region, we compared the performances of two PRSs computed by including the *APOE* SNP or not. Although *APOE* e4 SNP rs429358 is not in the list of 23 candidate SNPs, another SNP rs41289512 from the *APOE* region is part of the list. Given the linkage disequilibrium (LD) between rs429358 and rs41289512 ($r^2 = 0.214, D' = 0.967$), we note that rs41289512 may share some *APOE* e4 effect. With this observation, we constructed two different PRSs (PRS+ and PRS-) using summary statistics of AD candidate SNPs, and performed a comparative mediation study of amyloid imaging phenotypes for the following four genetic scores including two PRSs and two *APOE* SNPs:

- **PRS+**: PRS computed using all 23 AD candidate SNPs in Table 2
- **PRS-**: PRS computed using 22 AD candidate SNPs (without rs41289512) in Table 2
- **rs429358**: *APOE* e4 SNP, the best known AD genetic risk factor
- **rs41289512**: *APOE* SNP identified by the recent meta-GWAS study[3]

Mediation analysis

For each of the four genetic scores, we followed Baron and Kenny (1986)[18] to perform standard mediation analysis for identifying imaging phenotypes as potential disease mediators.

Let $y \in \{0, 1\}$ be the dependent variable which is a diagnostic phenotype (0:control, 1:case) in our study, x be the independent variable which is one of the four genetic scores, z be the covariates (age, sex and education), and $M = \{m_i\}, i \in (1, \ldots, 90)$ be the set of brain imaging mediators. Mediation analysis could be performed following the three steps listed as below.

Step 1: We use logistic regression to regress the diagnostic outcome y against the genetic score x, controlling for z:

$$logit(Pr(y = 1)) = \beta_{11}x + \beta_{12}z + \epsilon_1, \tag{2}$$

where the coefficient β_{11} should be significant (p-value < 0.05) to pass the first step.

Step 2: We use linear regression to regress the imaging mediator m_i against the genetic score x, controlling for z:

$$m_i = \beta_{21,i}x + \beta_{22,i}z + \epsilon_{2,i}, \tag{3}$$

Here the coefficient $\beta_{21,i}$ should be significant after correcting for multiple comparisons. We employ Bonferroni correction in this step, that is, the significant p-value threshold is $0.05/90 = 5.56\text{E-}04$.

Step 3: We use logistic regression to regress diagnostic outcome y against both the genetic score x and the mediator m_i, controlling for z:

$$logit(Pr(y = 1)) = \beta_{31,i}x + \beta_{32,i}m_i + \beta_{33,i}z + \epsilon_{3,i}. \tag{4}$$

Note step 3 is performed on only mediators passing the second step. We again employ Bonferroni correction for multiple comparisons, that is, correct for the number of mediators surviving step 2. The coefficient $\beta_{32,i}$ should be significant and $|\beta_{31,i}| < |\beta_{11}|$ namely an indirect effect is present between outcome y and PRS x mediated through m_i.

Table 3: Five diagnostic case-control groups included in mediation analysis.

Group	Control	Case	# of subjects
1	HC	AD+EMCI+LMCI	759
2	HC	AD	344
3	HC	LMCI	373
4	HC	EMCI	450
5	HC	EMCI+LMCI	619

Table 4: Pearson correlation coefficients among two PRSs and two studied SNPs.

PCC	PRS+	PRS-	rs41289512	rs429358
PRS+	1	0.572	0.831	0.438
PRS-	0.572	1	0.018	0.073
rs41289512	0.831	0.018	1	0.484
rs429358	0.438	0.073	0.484	1

We perform mediation analysis to explore the underlying association between each genetic score and diagnosis. Given that AD is a progressive disease where the dementia symptoms gradually worsen over a number of years, it is critical to identify the biomarkers which can predict the disease at early stage. Therefore, we perform mediation analysis on five case-control groups, which focus on different disease stages. The five grouping schemes are shown in Table 3.

Mediated effect comparison

As evidenced in the "Mediation Analysis" section, the total effect of a genetic score on the diagnostic outcome is $\beta_{11} = \beta_{31,i} + (\beta_{32,i} \cdot \beta_{21,i})$, where the $\beta_{31,i}$ is the direct effect and $\beta_{32,i} \cdot \beta_{21,i}$ is the indirect effect. Indirect effect, also called mediated effect, measures the amount of mediation, that is, the effect of the genetic score on outcome due to the imaging mediator.

In order to compare the mediated effects across different brain regions, we follow Breen et al. (2013)[19] to calculate relative magnitude of the indirect effect to the total effect. The proportion of the effect mediated by m_i is calculated as follows:

$$\text{prop\_mediation}(i) = \frac{\beta_{32,i} \cdot \beta_{21,i}}{\beta_{31,i} + \beta_{32,i} \cdot \beta_{21,i}}. \tag{5}$$

Results

Correlation among studied PRSs and SNPs

Four genetic scores, including two PRSs and two *APOE* SNPs, are compared in the mediation analysis to detect the additional effect of AD candidate SNPs beyond the well-known *APOE* SNPs. We first compute the Pearson correlation coefficients (PCCs) among the four genetic scores to evaluate their relationships (see Table 4). Even though the calculation of PRS+ involves the SNP rs41289512 that is in LD with *APOE* SNP rs429358, correlation between PRS+ and rs429358 is not very strong (PCC = 0.438), suggesting that PRS+ provides additional, novel information outside of the well-known *APOE* SNP rs429358. Further, we argue that from the moderately high correlation between PRS- and PRS+ (PCC = 0.572), our construction of the PRS- as a latent variable can provide novel information outside of the *APOE* region altogether.

Mediation analysis summary

For each genetic score, mediation analyses are performed across five different case-control grouping schemes. We summarize the results in Table 5, by listing the p-values and odds ratios (ORs) of the first step as well as the number of discovered mediators in each mediation analysis experiment.

From Table 5, PRS+, compared with other three genetic scores, obtains the largest number of mediators across different diagnostic case-control experiments, except for HC vs AD group where *APOE* SNP rs429358 identifies 75 mediators.

Table 5: Mediation analysis summary. For each experiment, the p-value and odds ratio of the first step in mediation analysis are listed, as well as the number of mediators discovered. The smallest p-value and largest number of mediators across case-control groups are shown in bold.

Group	Result	PRS+	PRS-	rs41289512	rs429358
HC vs MCI+AD (N = 759)	p-value	3.47E-04	2.20E-02	5.38E-03	**1.53E-07**
	OR	1.42	1.22	1.33	1.71
	# mediator	**68**	38	61	54
HC vs AD (N = 344)	p-value	7.08E-06	4.05E-03	2.49E-04	**7.09E-13**
	OR	1.71	1.4	1.52	2.68
	# mediator	16	0	0	**75**
HC vs LMCI (N = 373)	p-value	3.47E-03	6.52E-02	1.66E-02	**8.86E-06**
	OR	1.38	1.22	1.3	1.67
	# mediator	**33**	0	2	31
HC vs EMCI (N = 450)	p-value	4.48E-02	1.40E-01	1.72E-01	**2.08E-02**
	OR	1.24	1.16	1.16	1.28
	# mediator	**20**	0	0	11
HC vs MCI (N = 619)	p-value	7.00E-03	6.69E-02	4.28E-02	**3.03E-04**
	OR	1.3	1.18	1.23	1.44
	# mediator	**41**	0	39	27

This demonstrates the power of the collective effect of AD candidate SNPs for discovering additional brain imaging mediators than only *APOE* SNPs. SNP rs429358 outperforms other PRSs in HC vs AD group, possibly because of both the relatively small sample size (N = 344) and strong effect of rs429358 on AD. PRS- mediation analyses identify mediators from only the first diagnostic grouping scheme (i.e., HC vs MCI+AD), since it could not pass the first step for other grouping schemes. That is, in our data, PRS- is not significantly associated with each individual diagnosis.

Given the best performance of mediation analysis in the complete data (i.e., HC vs MCI+AD), we compare the mediators identified from different genetic scores. Figure 1(a) shows the intersections of mediators among PRSs on the HC vs MCI+AD group. Totally 68 unique ROIs are identified across the four experiments, 33 of which are commonly captured by mediation analyses of all four genetic scores. Mediation of PRS+ captures all 68 mediators, of which 3 mediators are not reported by any other PRSs. This additional findings from PRS+ mediation indicate the effectiveness of combination of AD candidate SNPs beyond *APOE* rs429358 for mediator detection. Although rs41289512 and *APOE* SNP rs429358 are in the LD, rs41289512 and PRS+ which includes the genetic effect of rs41289512, still capture additional 6 mediators than rs429358 and PRS-, suggesting the role of *APOE* region (containing additonal information on top of the best-known *APOE* e4 SNP) in the path from gene to diagnosis. Compared with rs429358, 5 ROIs are identified by PRS+, PRS- and rs41289512, further demonstrating the complex and underlying molecular mechanism of AD and the necessity of investigating the joint effect of genetic variants. To better illustrate the brain location of these interesting imaging mediators, we map the ROIs labeled in Figure 1(a) to the brain, and show them in Figure 1(b). We further discuss the detailed functions of identified imaging mediators in the next section.

Stagewise brain region indicators

It is critical and urgently required to predict AD in its early stage, which is also difficult due to the divergent and complex path from gene to disease. Our mediation analysis of imaging phenotypes could help understand the underlying molecular mechanism, by discovering intermediate brain measures which have more proximate biological link to genetic basis. Accordingly, the mediators identified from different disease diagnostic groups would help recognize disease stage-specific brain regions.

Figure 2 shows two heatmaps of results from PRS+ and rs429358 mediation analyses, presenting the proportions of indirect effect mediated by brain ROIs across experiments on different diagnostic groups, with brain ROIs exhibiting similar effect sizes clustered together. A total of 8 clusters are formed and illustrated by red boundary lines. Figure 3 shows the brain map of these clusters. These clusters are further mapped to brain structural regions according to their anatomical locations and shown in Table 6.

Overall, PRS+ which represents a more comprehensive genetic view of AD risk, presents more stable performance

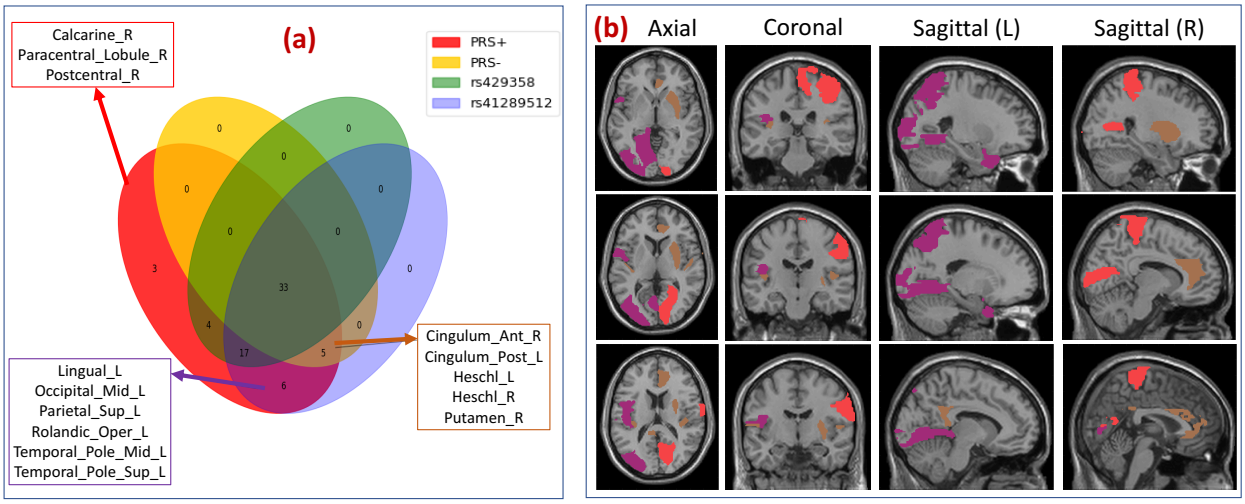

Figure 1: (a) Mediators discovered over different PRSs for HC vs. MCI+AD. (b) Brain mapping of regions discovered without rs429358

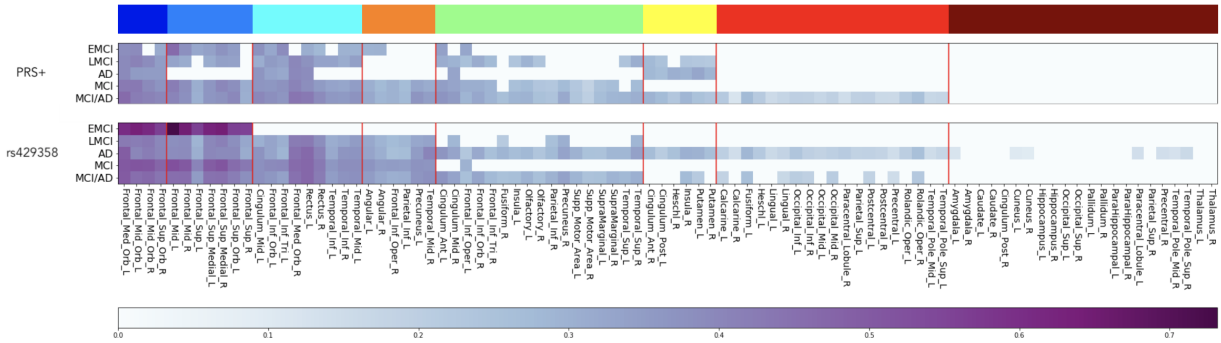

Figure 2: Heatmaps of indirect effect sizes of brain regions as mediator across mediation analyses on different diagnostic groups. Top panel shows the result of PRS+ mediation analysis and bottom panel shows the result of *APOE* rs429358. Color bar in the top row indicates brain region groups.

than only rs429358 on mediator discovery in the whole dataset (i.e., HC vs MCI+AD). While rs429358 performs more stably in HC vs AD group, mostly due to its strong effect on AD. From both PRS+ and rs429358 mediation results, regions included in the first two clusters are all located in the frontal lobe, and show strong mediation effects across all stages of AD. Some ROIs are not picked up in PRS+, possibly due to the limited sample size. PRS+ allows the discovery of more mediators for EMCI patients compared with using just rs429358, as shown in cluster 3 where most ROIs in PRS+ analysis show mediation effects in HC vs EMCI group while none are captured by rs429358. This indicates the sensitivity of comprised AD candidate SNPs for their collective effect on detection of early stage disease biomarkers. ROIs grouped in clusters 4 and 5 distribute in various brain regions, and present different mediation effects for different stages. For ROIs in cluster 4, both PRS+ and rs429358 detect their indirect effect in the HC vs MCI and HC vs MCI+AD groups, while only left and right angular gyrus are captured by PRS+ in HC vs EMCI group. This merit further examination on angular gyrus for its possible role as an early disease indicator.

Discussion

In this study, we have proposed to employ polygenic mediation analysis into imaging genetics of AD, to detect brain regions mediators for linking the complex path from genetics to diagnosis. The discovered imaging mediators intermediate the molecular modulation of AD candidate variants on disease, which may not be directly detected from

Table 6: Anatomical location of grouped brain regions from mediation analysis. For each cluster, the largest number of ROIs across brain structural regions is shown in bold.

Cluster	Subcortical	Frontal	Cingulate	Parietal	Temporal	Occipital	Insula	Sensory-Motor Cortex
1	0	**4**	0	0	0	0	0	0
2	0	**7**	0	0	0	0	0	0
3	0	**5**	1	0	3	0	0	0
4	0	1	0	**4**	1	0	0	0
5	0	**5**	2	4	**5**	0	1	0
6	**2**	0	**2**	0	1	0	1	0
7	0	2	0	1	**6**	**6**	0	4
8	**10**	0	1	1	4	4	0	2

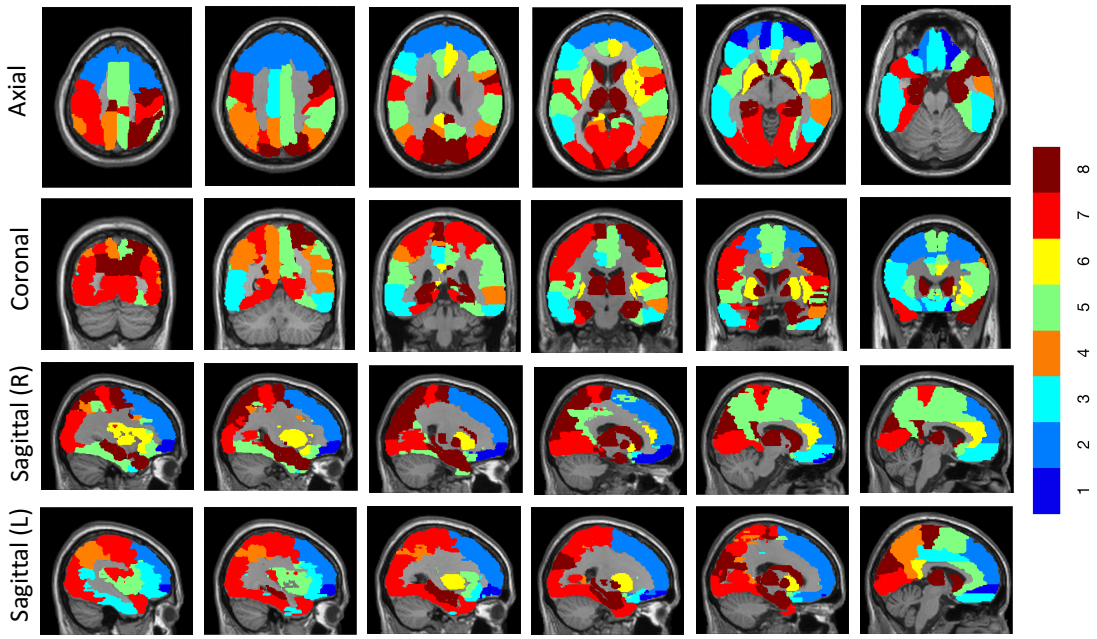

Figure 3: Brain mapping of clustered regions from our mediation analyses (see clustering results in Figure 2).

traditional SNP-outcome association analysis. We assess the performance of PRS comprised from multiple AD candidate SNPs on mediator detection, and compare with the best-known AD genetic risk factor in the *APOE* gene. To the best of our knowledge, this is among the first analysis in the study of AD for exploring mediation effect of imaging features underlying the PRS and disease. Our polygenic mediation analysis has identified multiple brain imaging mediators, a few of which present promises as early disease indicators.

We performed mediation analyses of two PRSs and two SNPs (i.e., PRS+, PRS-, rs41289512 and rs429358), among which PRS+ shows more power on mediator detection than the others, especially on the complete set of participants (N = 759; see Figure 1). In the HC vs MCI+AD study, PRS+ mediation model discovers three novel mediators undetected by other genetic scores, including right calcarine sulcus, right paracentral lobule and right postcentral gyrus. In addition, 6 mediators are detected by both PRS+ and rs41289512, and 5 mediators are detected by PRS+, PRS- and rs41289512. These novel findings indicate the cumulative effect of combined AD candidate SNPs beyond *APOE* e4 SNP rs429358 and may help provide more insights into the imaging genetic mediation mechanism of AD.

Based on the above observation, in the following, we focus on comparing the discoveries from PRS+ with those from rs429358 for different diagnostic groups in comparison with HC. From Figure 2, mediators detected by PRS+ and rs429358 illustrated both consistent and distinct patterns for different diagnostic groups across the 8 clusters. All 4

frontal lobe regions contained in cluster 1 show consistent intermediate effects across all stratified analyses in both PRS+ and rs429358 mediation models, suggesting the involvements of these four regions as intermediate modulators from early to late stages of AD. In cluster 2, regions are located in the frontal lobe and consistently detected except the PRS+ on HC vs AD group, possibly because of the modest sample size (N = 344) and the small PRS effect size.

From clusters 3 and 4, no regions are detected by rs429358 as mediators in the EMCI group, while PRS+ effects in EMCI go through most cluster 3 regions as well as left and right angular gyri of cluster 4. This suggests that in early stage PRS+ may start to affect brain mechanisms in these specific regions, while the effect of rs429358 appears at late stage. These regions can help serve as early indicators and provide valuable information for tracking disease progression from very early stage. As mentioned above, the 3 novel detected mediators by PRS+ in the HC vs MCI+AD group are located in clusters 6 and 7. These regions also exist in the rs429358 mediation result on HC vs AD group, while are not identified from the whole dataset, suggesting the differential effects of PRS+ and *APOE* on these regions. That is, these regions may have implicit effects in the paths from both PRS+ and *APOE* to disease while in different stages.

From clusters 1 to 7, mediators detected by *APOE* are all from HC vs AD comparison, while mediators detected by PRS+ are all from HC vs MCI+AD comparison. This confirms the strong effect of *APOE* rs429358 in late AD, which however has not exhibited in the early stage. On the other hand, PRS+ may not have comparable effect on disease as rs429358 especially given small sample size, but it conserves the ability of mediating biomarkers through various stages given its joint effect from multiple candidate variants. This suggests the power of the collective effect of SNPs which are not individually significant while cumulatively affecting disease.

Conclusions

In conclusion, we performed PRS mediation analyses of Alzheimer's disease for detecting intermediate imaging phenotypes to bridge the gap between genetics and disease, by discovering the implicit effects of AD risk variants. We discovered amyloid imaging mediators using PRS comprised from AD candidate SNPs, and compared its effectiveness with the well-known *APOE* SNP rs429358. As it is critical to predict the disease in early stage, we further evaluated the power for early disease indicator discovery, where PRS outperformed the *APOE* SNP alone, showing the promise of PRS on biomarker detection. This work can be further expanded towards several future directions. For example, different strategies can be used to construct the PRS, like threshold-based approaches which evaluate the performance of different number of top SNPs from GWAS results. Another direction is to apply PRS mediation analysis to multi-modal imaging data for understanding genetic mechanisms from a more comprehensive perspective.

Acknowledgements

This work was supported in part by NIH R01 EB022574, RF1 AG063481, R01 LM013463, U19 AG024904, R01 AG019771, and P30 AG010133; and NSF IIS 1837964.

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.

References

1. J. C. Lambert et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat. Genet.*, 45(12):1452–1458, 2013.
2. L. Shen et al. Genetic analysis of quantitative phenotypes in AD and MCI: imaging, cognition and biomarkers. *Brain Imaging Behav.*, 8(2):183–207, 2014.
3. Iris E. Jansen, Jeanne E. Savage, et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nature Genetics*, 51(3):404–413, 2019.
4. A.R. Hariri, E.M. Drabant, and D.R. Weinberger. Imaging genetics: Perspectives from studies of genetically driven variation in serotonin function and corticolimbic affective processing. *Biological Psychiatry*, 59(10):888–897, 2006.
5. Jr. Jack, C. R., D. A. Bennett, K. Blennow, M. C. Carrillo, H. H. Feldman, G. B. Frisoni, H. Hampel, W. J. Jagust, K. A. Johnson, D. S. Knopman, R. C. Petersen, P. Scheltens, R. A. Sperling, and B. Dubois. A/T/N: An unbiased descriptive classification scheme for Alzheimer disease biomarkers. *Neurology*, 87(5):539–47, 2016.
6. L. Shen and P. M. Thompson. Brain imaging genomics: Integrated analysis and machine learning. *Proceedings of the IEEE*, 108(1):125–162, Jan 2020.
7. Xuan Bi, Liuqing Yang, Tengfei Li, Baisong Wang, Hongtu Zhu, et al. Genome-wide mediation analysis of psychiatric and cognitive traits through imaging phenotypes. *Human Brain Mapping*, 38(8):4088–4097, 2017.
8. Andrew J. Saykin, Li Shen, Xiaohui Yao, Sungeun Kim, Kwangsik Nho, et al. Genetic studies of quantitative MCI and AD phenotypes in ADNI: Progress, opportunities, and plans. *Alzheimers Dement.*, 11(7):792–814, 2015.
9. F. Dudbridge. Power and predictive accuracy of polygenic risk scores. *PLoS Genet*, 9(3):e1003348, 2013.
10. D. Chasioti, J. Yan, K. Nho, and A. J. Saykin. Progress in polygenic composite scores in alzheimer's and other complex diseases. *Trends Genet*, 35(5):371–382, 2019.
11. Irene López González, Paula Garcia-Esparcia, Franc Llorens, and Isidre Ferrer. Genetic and transcriptomic profiles of inflammation in neurodegenerative diseases: Alzheimer, Parkinson, Creutzfeldt-Jakob and Tauopathies. *International Journal of Molecular Sciences*, 17(2):206, 2016.
12. Vijay K Ramanan, Shannon L. Risacher, Kwangsik Nho, Sungeun Kim, Li Shen, Brenna C. McDonald, et al. GWAS of longitudinal amyloid accumulation on 18F-florbetapir PET in Alzheimer's disease implicates microglial activation gene IL1RAP. *Brain*, 138(10):3076–3088, 2015.
13. Michael W Weiner, Dallas P Veitch, Paul S Aisen, Laurel A Beckett, Nigel J Cairns, Robert C Green, Danielle Harvey, Clifford R Jack, John C Morris, et al. Recent publications from the Alzheimer's disease neuroimaging initiative: Reviewing progress toward improved ad clinical trials. *Alzheimer's & Dementia*, 13(4):e1–e85, 2017.
14. Xiaohui Yao, Jingwen Yan, et al. Two-dimensional enrichment analysis for mining high-level imaging genetic associations. *Brain Informatics*, 4(1):27–37, Mar 2017.
15. Xiaohui Yao, Shannon Risacher, Kwangsik Nho, Andrew Saykin, Ze Wang, and Li Shen. Targeted genetic analysis of cerebral blood flow imaging phenotypes implicates the *INPP5D* gene. *Neurobiol Aging*, 81:213–221, 2019.
16. Andrew J. Saykin, Li Shen, et al. Alzheimer's Disease Neuroimaging Initiative biomarkers as quantitative phenotypes: Genetics core aims, progress, and plans. *Alzheimer's and Dementia*, 6(3):265–273, 2010.
17. S. Purcell et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.*, 81(3):559–75, 2007.
18. Reuben M. Baron and David A. Kenny. The Moderator-Mediator Variable Distinction in Social Psychological Research. Conceptual, Strategic, and Statistical Considerations. *J Pers Soc Psychol.*, 51(6):1173–1182, 1986.
19. Richard Breen, Kristian Bernt Karlson, and Anders Holm. Total, direct, and indirect effects in logit and probit models. *Sociological Methods & Research*, 42(2):164–191, 2013.