Identifying Subtypes in Alzheimer's Disease

Stephanie Bi[1], Li Shen[2], Jingxuan Bao[2]

[1]Department of Computer and Information Science, School of Engineering and Applied Science, University of Pennsylvania, Philadelphia, PA, USA
[2]Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

Introduction

Background

- Key biomarkers of Alzheimer's disease (AD) can be used to measure the progression of the illness within a particular patient.
- One particular area of interest has been the subtyping of such biomarkers or, in other words, identifying a threshold such that all patients who meet such threshold can be determined to be at high risk for the disease.
- Subtyping allows for early detection of AD and possible treatment intervention.

Objective

- We want to define subtypes for 16 key biomarker candidates of AD and examine whether proposed subtypes produce significant genetic (SNP) bases.

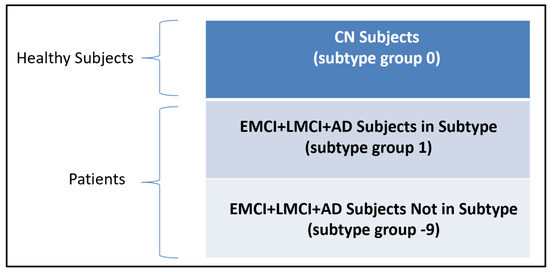


Figure 1. *Divided groups of subjects after subtyping.*

Materials

- Alzheimer's Disease Neuroimaging Initiative (ADNI) data: There were a total of 1576 participants and 565,373 SNPs assessed for significant associations.
- Biomarker candidates: AV45, FDG, PTAU, TAU, ABETA, CDRSB, ADAS13, MMSE, RAVLT.learning, FAQ, Ventricles, Hippocampus, WholeBrain, Entorhinal, Fusiform, and MidTemp
- Covariates: Patient age, gender, education, and subtype group (0, 1, or -9) were used.

Age	Sex (Male/Female)	Education(Year)
73.87 ± 7.28	889/687	15.85 ± 2.87

Figure 2. *Participants' demographic information*

GWAS analysis:

- We used a p-value significance threshold of $5*10^{-8}$ to determine whether a significant SNP association existed after running PLINK.

Methodology

Histogram analysis of CN (Cognitively normal), EMCI (Early mild cognitive impairment), LMCI (Late mild cognitive impairment), and AD data

Genome-wide association studies (GWAS) analysis on subtyped data comparing CN to EMCI+LMCI+AD patients within defined subtype

If GWAS returns no significant SNP associations, return to Step 1 and select new threshold.

Statistical analysis (pairwise t-test)

Results

- Various thresholds were assessed as possible subtype thresholds for each biomarker.
  - For each biomarker, subjects were split into the three groups and GWAS analysis was used to determine whether each subtype yielded significant genetic (SNP) associations.
- We used one subtype to define other biomarkers and created violin plots of all 16 biomarkers and their controls.
  - A pairwise t-test was performed between each biomarker's subtype and control to determine whether the subtype used to define the groups produced significant groupings in other biomarkers.

Histogram Results

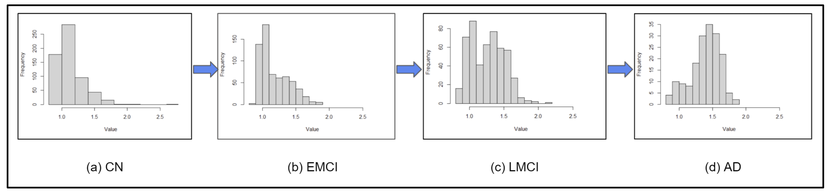


Figure 3. *Example histograms for AV45*

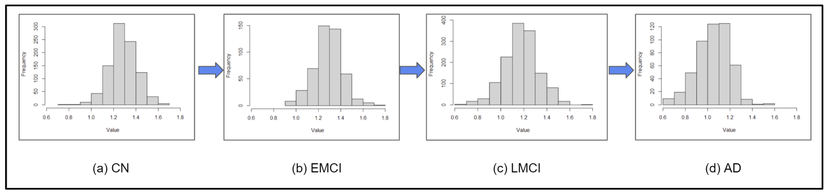


Figure 4. *Example histograms for FDG*

AV45	FDG	PTAU	TAU	ABETA	CDRSB	ADAS13	MMSE	RAVLT.learning	FAQ	Ventricles	Hippocampus	WholeBrain	Entorhinal	Fusiform	MidTemp
L	S	L	L	S	L	L	S	S	L	L	S	S	S	S	S

Figure 5. *Directionality results (L: larger values signify AD progression, S: smaller values signify AD progression)*

GWAS Results

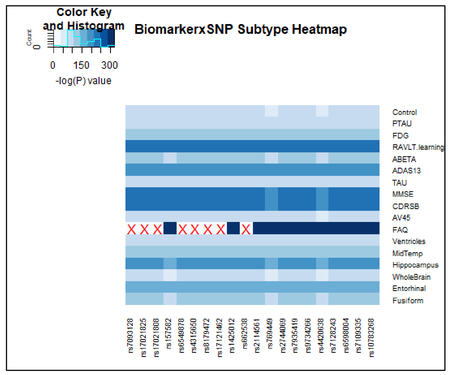


Figure 6. *Heatmap displays associations between biomarkers and SNPs received from PLINK analysis. Red X's represent the most significant associations. X-axis represents SNPs; y-axis represents biomarkers*

AV45	FDG	PTAU	TAU	ABETA	CDRSB	ADAS13	MMSE	RAVLT.learning	FAQ	Ventricles	Hippocampus	WholeBrain	Entorhinal	Fusiform	MidTemp
>1.4	<1.1	>50	>500	<850	>9	>40	<27	<5	>15	>80000	<6000	<1000000	<3000	<16000	<18000

Figure 7. *Subtyping results of all biomarkers*

Results (cont.)

Pairwise T-test Results

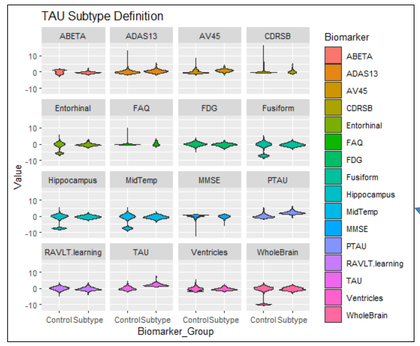


Figure 9. *Enlarged violin plot for TAU*

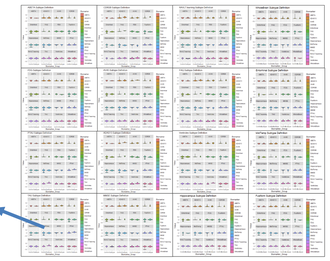
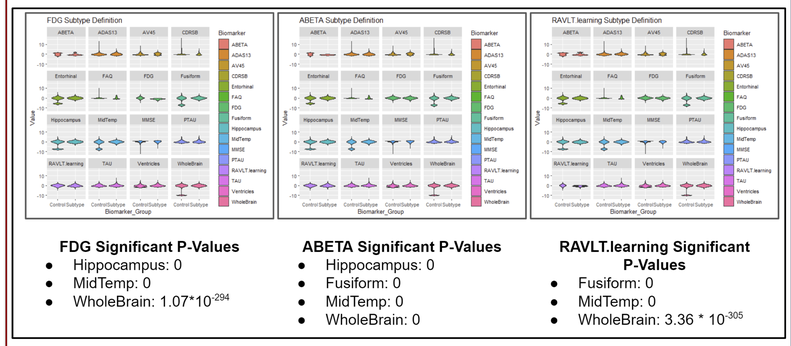


Figure 8. *Violin plots with normalized data of all biomarkers*



FDG Significant P-Values
- Hippocampus: 0
- MidTemp: 0
- WholeBrain: $1.07*10^{-294}$

ABETA Significant P-Values
- Hippocampus: 0
- Fusiform: 0
- MidTemp: 0
- WholeBrain: 0

RAVLT.learning Significant P-Values
- Fusiform: 0
- MidTemp: 0
- WholeBrain: $3.36 * 10^{-305}$

Figure 10. *Most significant t-test results for possible associations between various biomarkers*

Conclusion

- Out of the 16 biomarker candidates subtyped and assessed, we determined that the threshold of 15 for FAQ yielded the most significant genetic associations and could be further explored as a possible indication of AD.
- Our findings also indicate that a few biomarkers could possibly be associated with others and used to indicate the development of AD, such as that between FDG and MidTemp or ABETA and WholeBrain.
- Having focused on GWAS and biomarker evaluation in this project, in the future, we could move to analyze survival outcome as well.
- We could also use other data to back our findings in this study and use other strategies such as ones involving machine learning to subtype patients.

Acknowledgements

- Supported by the Shen Lab at the Perelman School of Medicine at the University of Pennsylvania and the Penn Undergraduate Research Mentoring (PURM) program.
- Full ADNI Acknowledgement is available at: https://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.