GWAS on Alzheimer's Disease

Haoming Zhang, Zhuoyuan Ren, Haoshu Qin

May 2020

Abstract

Alzheimer's Disease (AD) is a progressive disease that could cause physical and psychological damages, including memory loss, personality alteration, disorientation, etc. Recently, some studies have identified Apolipoprotein E (APOE) as the most gene associated with Late-Onset Alzheimer's Disease. Scientists are now working to discover more non-APOE genes that may be associated. In this study, we tested nine non-APOE genes reported in previous studies on the dataset from University of Miami, including 176 cases and 188 controls. We also conducted association tests on the dataset and compared it with the result of the meta-analysis on three different data sets. Nominal significant associations ($P < 0.05$) were observed on two genes (CR1 and PICALM) out of the nine we tested on. We also found 3 genes (CSNK1E, CACNG2, LINGO2) that showed significant associations during the hypothesis test. However, none of these 3 genes were accepted in the meta-analysis.

1 Introduction

1.1 Alzheimer's Disease

Alzheimer's disease (AD) is a degenerative brain disease and the most common cause of dementia. Alzheimer's disease represents a major public health concern and has been identified as a research priority [14]. Patients suffer from memory loss, challenges in panning or solving problems, confusion with time or place, changes in mood and personality, increased anxiety, agitation and sleep disturbances. An estimated 5.4 million Americans have Alzheimer's disease. By the mid-century, the number of people living with Alzheimer's disease in the United States is projected to grow to 13.8 million[2]. There are three common kinds of Alzheimer's disease, Early-onset Alzheimer's disease (EOAD), Late-onset Alzheimer's disease (LOAD), and Familial Alzheimer's disease (FAD). EOAD and LOAD are differentiated as the patients of the latter are over 65 years older when they are diagnosed. In this study, we will look at cases of LOAD which takes up to more than 90% of all AD cases in order to potentially benefit a wider range of population [5].

The cause of Alzheimer's disease is poorly understood. About 70% of the risk is believed to be inherited from a person's parents with many genes usually involved. Other risk factors include a history of head injuries, depression, and hypertension.[14] The goal of our study is to identify potential risk genes or loci that have significant associations with LOAD.

1.2 Previous GWAS

Historically, Apolipoprotein E(APOE) is widely accepted as a major genetic risk determinant of LOAD and therefore is not included as one of the risk genes to be studied by many studies anymore [15].

In GWAS study [7], the researchers used a GWAS data set from the University of Pittsburgh (1291 cases and 938 controls) to examine in detail implicated nine new regions (CR1, BIN1, CLU, PICALM, MS4A4/MS4A6E, CD2AP, CD33, EPHA1 and ABCA7) with Late-Onset Alzheimer's disease (AD) risk, and also performed a meta-analysis utilizing the top 1% GWAS single-nucleotide polymorphisms (SNPs) with $P < 0.01$ along with four independent data sets (2727 cases and 3336 controls) for these SNPs in an effort to identify new AD loci.

Nominal significant associations ($P < 0.05$) were observed either within or adjacent to five genes (PICALM, BIN1, ABCA7, MS4A4/MS4A6E and EPHA1), significant signals were observed 69–180 kb outside of the remaining four genes (CD33, CLU, CD2AP and CR1). Meta-analysis on the top 1% SNPs showed association in the PPP1R3B gene (top SNP rs3848140 with P=3.05E–07). The association of this SNP with AD risk was consistent in all five data sets with a meta-analysis odds ratio of 2.43. [7]

We were inspired by this study and hoped to replicate the association analysis of these nine regions on another data set to see if the results remain consistent across different data sets. Also, we tried to detect other potential genes other than those nine regions. Therefore, our hypothesis is that the conclusions about the nine genes (CR1, BIN1, CLU, PICALM, MS4A4/MS4A6E, CD2AP, CD33, EPHA1 and ABCA7) in [7] remain consistent in our different data set and there are no new genes associated with LOAD would be found.

1.3 Data

1.3.1 Data sets

Our data set comes from Dr. Amanda J. Myers at University of Miami School of Medicine. The data set includes genotyping calls from 502,627 SNPs on the 364 samples (191 males and 173 females, 176 are cases and 188 are controls). The control cohort and case cohort were obtained from 20 National Alzheimer's Coordinating Center (NACC) brain banks and from the Miami Brain Bank. The criteria for inclusion were as follows: self-defined ethnicity of European descent,

neuropathologically confirmed LOAD or no neuropathology present, and age of death greater than 65. We restricted samples to only of European ethnicity because allele-frequency differences between groups with different ethnicity might create bias in the results.[6].

For meta-analysis, we found three additional data sets. Data set [1] includes 3,444 controls and 653 controls; Data set [8] includes 27,696 cases of maternal AD (260,980 controls) and 14,338 cases of paternal AD (245,941 controls); Data set [10] contains 17,008 Alzheimer's disease cases and 37,154 controls. The samples on the three data sets are all white European and the diseases are identified as LOAD.

1.3.2 Quality Control

For quality control, we first filtered out 16555 variants that failed a Hardy-Weinberg equilibrium exact test at p value 0.05. Next, we filtered out 2621 variants due to their minor allele frequency below threshold 0.01; Following that, we filtered out 4411 variants with missing call rates exceeding 10 percent per variant.; Next, 0 sample were filtered out with missing call rates exceeding 10 percent per sample. Furthermore, we filtered out variant from outside of chromosome 1 to chromosome 22, because we only care about the 22 pairs of autosomes but not the two sex chromosomes, X and Y; In addition, we checked if the sex assignment in the input file aligned with those chromosomal data imputed from X chromosome inbreeding coefficients. 380157 variants and 364 people passed; At last, we filtered for cryptic relatedness using –min .20 to remove lines with PI_HAT values below 0.20. 0 sample were removed and 348967 variants in 364 subjects passed. The final total genotyping rate was 0.976.

1.3.3 Exploratory Data Analysis

For exploratory data analysis, we ran principal component analysis in Plink[9]. Then, we created a matrix plot (Figure 1) to compare each one of the first five principal components against themselves and the other four. In the first row where we plotted PC1 on Y axis and PC1-5 on X axis, we can observe two to three separate clusters formed in these four graphs. However, no separated clusters are found in graphs where we plot PC2, PC3, PC4 against other PCs. So only PC1 seems to separate the subjects into clusters, so we chose PC1 as a covariate (independent variable) for running logistic regression on the data.

Figure 2 shows the scree plot of top 20 PCs. The x-axis shows each PC and the y-axis show the eigenvalue of each PC, the higher the engenvalue, the more associative the PC is. From our scree plot, we can see that PC1 one has the highest eigenvalue compare to all the rest of the PCs, combining with the matrix plot from above, we will choose only PC1 as a covariate (independent variable) for running logistic regression on the data later.

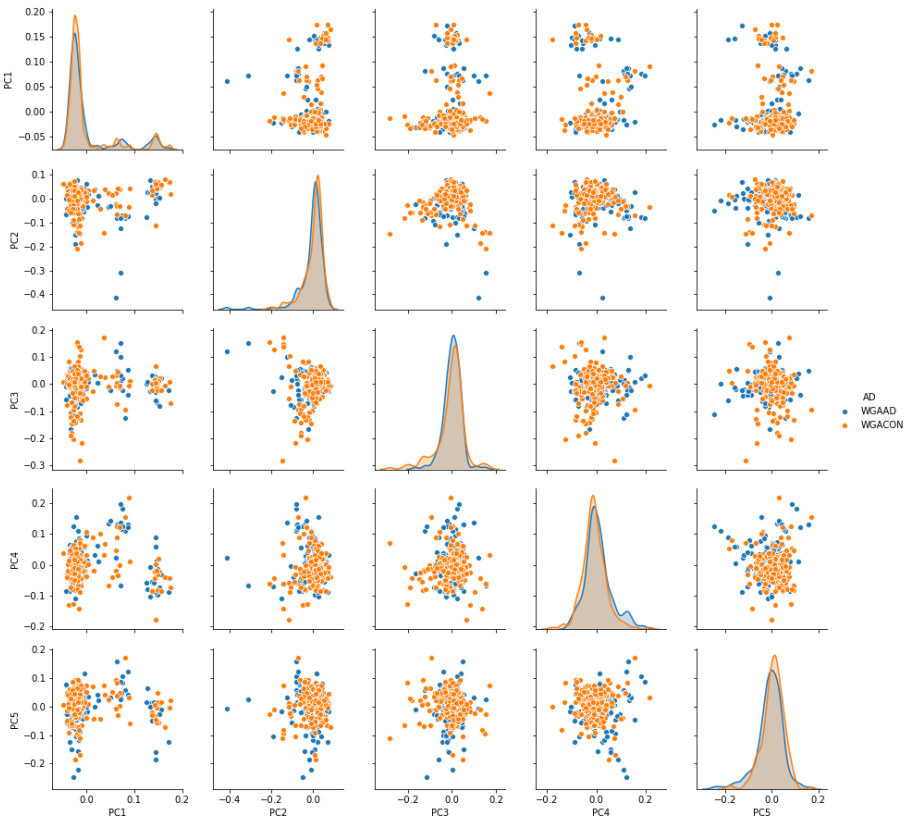

Figure 1: PCA Scatter Matrix

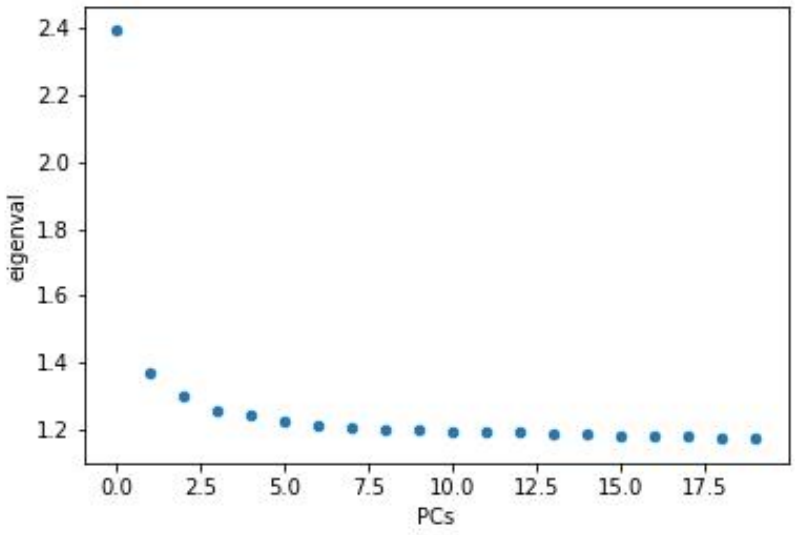

Figure 2: Scree Plot of PC

2 Methods

2.1 GWAS Analysis

We conducted association tests for each SNP in our data set using PC1, age, sex as covariates, with null hypothesis that the genetic variation at the locus has no association with LOAD. We get p-values after the test, which, statistically, represent the probabilities that we get the observed results assuming the null hypothesis is correct. The lower the p-value is , the more likely the SNP would be associated with the disease. Since we are doing multiple independent tests, we hope to control the family-wise error rate (FWER) at 0.05, which is the probability of having at least one false discoveries. According to [3], 5×10^{-8} is widely used for GWAS on White European population as the threshold for significant association, so we followed the tradition here as well. We used the following tags:

a. –adjust: Report basic multiple-testing adjustments for association test p-values

b. –allow-no-sex: Do not force ambiguous-sex phenotype to missing

c. –ci .95: Report confidence intervals for odds ratios

d. –logistic: Multi-covariate association analysis on a case/control phenotype

We explored our result from the association tests with the help of two kinds

of plots: a Manhattan plot and regional association plots. Figure 5 displays the p-value of SNPs on a large scale, the x-axis consists of all 22 chromosomes and the y-axis include the negative log of p-values. The plot gives an overview of all the SNPs based on their chromosomal locations, and can help us quickly identify which chromosome is more likely to be associated with LOAD.

Figure 4 to Figure 12 are regional association plots focusing on specific genes. It shows the p-values of SNPs at specific genetic locations on the chosen chromosomes (500 kb from both side of the gens). The spread of x-axis is normally measured in kilo-base pair or mega-base pair.

2.2 Meta Analysis

One medical problem is usually studied by different groups of researchers. To determine the correctness of a study, meta-analysis is widely adopted to compare with previous studies on the same problem. Meta-analysis is a quantitative, formal, epidemiological study design used to systematically assess previous research studies to derive conclusions about that body of research. It brings out a general and precise estimate of the effect of the risk factor for disease.

Our study consists of two stages. In stage 1, we conducted GWAS analysis (association test) on the GWAS data set from the University of Miami in an effort to identify new SNPs that might be associated with AD. For stage 2, we conducted a fixed-effects meta-analysis on three different Alzheimer's Disease related data sets to verify the correctness of the results from stage 1.

3 Results

3.1 Association Analysis in 9 none-APOE regions

We ran logistic regression and output the genetic loci and corresponding P values. As in Figure 3, each row of the table contains the SNP, chromosome it belongs to, original P value associated with LOAD and etc. We sorted the table in the ascending order of P value in order to observe SNPs that are most likely to be associated with LOAD.

Regional association plot shows a close up of P-values of specific genes and its surroundings. We created a regional association plot for each one of the nine non-APOE genes (CR1, BIN1, CLU, PICALM, MS4A4/MS4A6E, CD2AP, CD33, EPHA1 and ABCA7).

In the CR1 gene and its surroundings (+/- 5000000 base pairs) on chromosome 1 (Figure 8), 61 SNPs showed significant association ($P < 0.05$). Out of these significant SNPs, two SNPs are located on the gene: rs7514606 ($P = 0.04243$) and rs6540648 ($P = 0.04565$), 39 SNPs are located on the left side of

	CHR	SNP	BP	A1	TEST	NMISS	OR	SE	L95	U95	STAT	P
658132	19	rs429358	50103781	C	ADD	364	6.2410	0.2173	4.0760	9.5550	8.426	3.566000e-17
658134	19	rs4420638	50114786	C	ADD	364	4.4380	0.1923	3.0450	6.4700	7.750	9.171000e-15
694702	22	rs867198	37033106	T	ADD	345	5.4450	0.3194	2.9110	10.1800	5.306	1.122000e-07
276384	6	rs3011823	83208939	G	ADD	340	10.2800	0.4508	4.2490	24.8700	5.169	2.356000e-07
68958	2	rs9309095	43147102	T	ADD	335	2.1800	0.1662	1.5740	3.0190	4.688	2.755000e-06
694336	22	rs4821510	35327956	T	ADD	361	0.3785	0.2143	0.2487	0.5761	-4.533	5.810000e-06
457216	11	rs1155331	22342766	T	ADD	349	2.3650	0.1924	1.6220	3.4480	4.474	7.664000e-06
515662	12	rs11114028	107598413	T	ADD	358	2.0770	0.1696	1.4890	2.8960	4.309	1.642000e-05
386418	9	rs7043927	28884989	C	ADD	349	0.3948	0.2172	0.2579	0.6043	-4.280	1.873000e-05
575622	15	rs1393404	35933784	T	ADD	340	0.5089	0.1591	0.3726	0.6952	-4.246	2.180000e-05
366008	8	rs16883408	113400729	C	ADD	342	9.7670	0.5437	3.3650	28.3500	4.191	2.772000e-05
317386	7	rs7800867	69831614	T	ADD	331	4.1540	0.3425	2.1230	8.1280	4.158	3.209000e-05
457214	11	rs11026531	22335760	T	ADD	356	2.2890	0.2011	1.5430	3.3950	4.117	3.839000e-05
397544	9	rs7847449	97631463	A	ADD	362	0.5144	0.1626	0.3740	0.7074	-4.090	4.316000e-05

Figure 3: Genetic loci associated with the risk of AD

the gene, and 20 SNPs are located on the right side of the gene.

In the BIN1 gene and its surroundings (+/- 5,000,000 base pairs) on chromosome 2 (Figure 9), 49 SNPs showed significant association ($P < 0.05$). Out of these significant SNPs, no SNPs are located on the gene, 37 SNPs are located on the left side of the gene, and 12 SNPs are located on the right side of the gene.

In the CD2AP gene and its surroundings (+/- 5,000,000 base pairs) on chromosome 6 (Figure 10), 57 SNPs showed significant association ($P < 0.05$). Out of these significant SNPs, no SNPs are located on the gene, 31 SNPs are located on the left side of the gene, and 27 SNPs are located on the right side of the gene.

In the EPHA1 gene and its surroundings (+/- 5,000,000 base pairs) on chromosome 7 (Figure 11), 61 SNPs showed significant association ($P < 0.05$). Out of these significant SNPs, no SNPs are located on the gene, 37 SNPs are located on the left side of the gene, and 24 SNPs are located on the right side of the gene.

In the CLU gene and its surroundings (+/- 5,000,000 base pairs) on chromosome 8 (Figure 12), 91 SNPs showed significant association ($P < 0.05$). Out of these significant SNPs, no SNPs are located on the gene, 36 SNPs are located on the left side of the gene, and 55 SNPs are located on the right side of the gene.

In the MS4A4/MS4A46E gene and its surroundings (+/- 5,000,000 base pairs) on chromosome 11 (Figure 13), 81 SNPs showed significant association ($P < 0.05$). Out of these significant SNPs, no SNPs are located on the gene, 25 SNPs are located on the left side of the gene, and 56 SNPs are located on the right

side of the gene.

In the PICALM gene and its surroundings (+/- 5,000,000 base pairs) on chromosome 11 (Figure 14), 62 SNPs showed significant association ($P < 0.05$). Out of these significant SNPs, 3 SNPs are located on the gene: rs1938936 ($P = 0.04430$), rs10898524 ($P = 0.04447$), and rs7935611 ($P = 0.04535$), 35 SNPs are located on the left side of the gene, and 24 SNPs are located on the right side of the gene.

In the ABCA7 gene and its surroundings (+/- 5,000,000 base pairs) on chromosome 19 (Figure 15), 15 SNPs showed significant association ($P < 0.05$). Out of these significant SNPs, no SNPs are located on the gene, 12 SNPs are located on the left side of the gene, and 3 SNPs are located on the right side of the gene.

In the CD33 gene and its surroundings (+/- 5,000,000 base pairs) on chromosome 19 (Figure 16), 34 SNPs showed significant association ($P < 0.05$). Out of these significant SNPs, no SNPs are located on the gene, 17 SNPs are located on the left side of the gene, and 17 SNPs are located on the right side of the gene.

3.2 GWAS Analysis

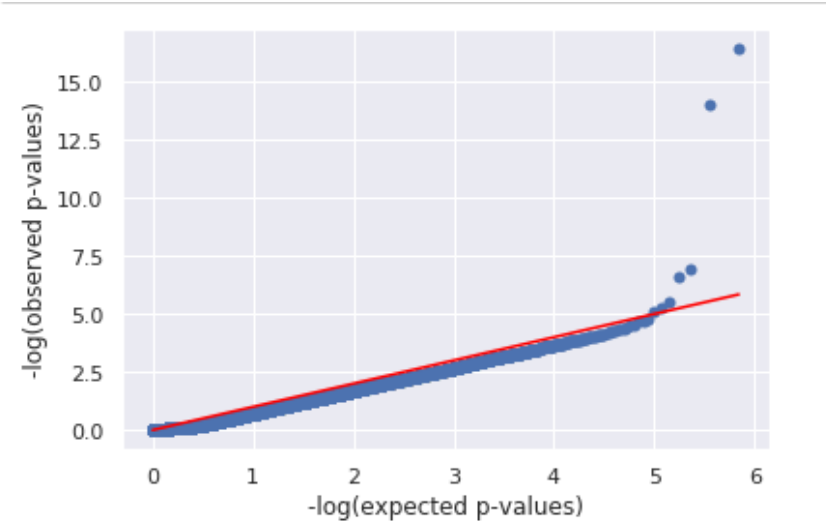

Figure 4: QQ Plot

QQ-plot stands for the quantile-quantile plot which compares the observed p-value versus expected p-value under the null hypothesis that there is no association between the variants and LOAD. In Figure 4, we plotted log of observed

p-value on y-axis against the log of expected p-value on x-axis. If all the dots follow a straight line it means that the observed p-value follows a uniform distribution and there is no causal polymorphism. We can observe a tail departing from the straight line on the right side of the QQ plot, so there are some SNPs with significant p-values that suggest potential LOAD associations.

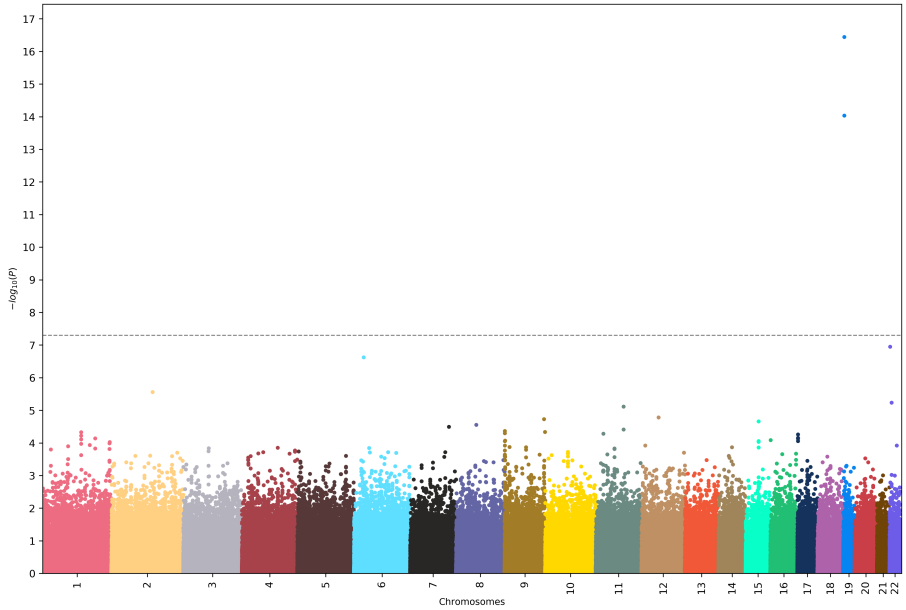

Figure 5: Manhattan Plot

Manhattan plot in Figure 5 shows the association of LOAD to each one of the 22 chromosomes. Each data point represents one SNP. We have negative log of p-values on the y-axis and chromosomes number on the x-axis. The higher the value on the y-axis, the more correlated that the corresponding SNP is with LOAD. We can observe a couple of SNPs with extreme p-values on chromosome 19, implying that chromosome 19 should contain genes that are highly correlated with LOAD. Other than chromosome 19, chromosome 22, 6 also contain SNPs with relatively small p-values.

We incorporated three additional data sets to produce the meta-analysis result shown in Figure 6. We sorted the result by the p-value of SNPs in ascending order to observe top significant SNPs that might be correlated with LOAD.

4 Discussion

4.1 Significant SNPs

Gene	Kamboh 2012[7]	Ours
CR1	No	Yes
BIN1	Yes	No
CD2AP	No	No
EPHA1	Yes	No
CLU	No	No
PICALM	Yes	Yes
MS4A4/MS4A6E	Yes	No
CD33	No	No
ABCA7	Yes	No

Table 1: Comparison between Paper [7] and Our Results on Each Gene's Association with LOAD

SNP	Chromosome	Gene	Related Disease
rs429358	19	APOE	AD
rs4420638	19	APOC1	AD
rs867198	22	CSNK1E	cerebral disease
rs3011823	6	TPBG/LOC105377876	kidney and bladder cancer
rs9309095	2	LINC01819	
rs4821510	22	CACNG2	mental retardation
rs1155331	11	SLC17A6	deafness
rs11114028	12	CORO1C	
rs7043927	9	LINGO2	parkinson, late-onset
rs1393404	15	TMCO5A	spastic paraplegia

Table 2: Significant SNPs and Related Diseases

In Table 1 we make comparisons between our association analysis results about the 9 non-APOE regions (CR1, BIN1, CLU, PICALM, MS4A4/MS4A6E, CD2AP, CD33, EPHA1 andABCA7) versus the results in [7]. We reached same conclusions that gene (PICALM) could be associative with LOAD while genes (CD2AP, CLU, CD33) are not associative to LOAD. Differently, we concluded that gene (CR1) should be associative with LOAD and genes (BIN1, EPHA1, MS4A4/MS4A63, ABCA7) should not be while the previous GWAS [7] held opposite views.

Table 2 displays the top 10 most significant SNPs from Figure 3, as well as the chromosomes, gene names and related diseases. We found SNPs (rs429358, rs4420638, rs867198, rs7043927, rs4821510) associated with certain neural dis-

eases. These SNPs corresponds to genese APOE, APOC1, CSNK1E, CACNG2, and LINGO2 respectively. APOE/APOC1 genes on chromosome 19 are commonly known genes to be associated with AD [4]; CSNK1E gene on chromosome 22 is associated with severe brain disorder, which can be related to AD [12]; CACNG2 on chromosome 22 is associated with mental retardation, which is one of the symptoms of LOAD [11]; LINGO2 on chromosome 9 is associated with Parkinson Disease, which is also a neuro-disease that have similar symptoms as LOAD [13]. Half of our significant SNPs showed association to either LOAD or related diseases, this suggests our findings to be significant and credible.

We looked up these five significant SNPs in the meta analysis result and found their p-values shown in Figure 7. We can see that out of our ten significant SNPs, only two of them (rs4420638, corresponds to APOC1 gene) and (rs429358, corresponds to APOE gene) appear to be significant. The other 3 genes (rs867198, rs7043927, rs4821510) do not appear to be associative with LOAD across the three additional data sets we used in meta analysis.

4.2 Limitations

The main limitation of our project was the relatively small size of the data set. We planed to use two data sets containing thousands of subjects, but only managed to gain access to the current data set containing a few hundreds of subjects. This could let the variances in our results relatively high and therefore less accurate to the universal truth.

Also, the three additional data sets' summary statistics that we collected might not have been obtained in a very consistent pipeline. Most ideally, three additional data sets and our main data set should all have been processed in identical quality control process and association analysis. However, we could not achieve that since we only have the summary statistics instead of original data of those three data sets. The inconsistency of quality control across different data set could affect the accuracy of the result of our meta-analysis.

Lastly, we hoped to study genes related to LOAD for all races of subjects, but as allele-frequency differences between groups with different ethnicity might create bias in the results and current data sets are predominantly of white European subjects, our study results are only helpful to a subset of patients around the world.

5 Future Plan

Our result can give valuable insights for medical professionals and researchers. By providing specific locations associated with AD, researchers can narrow down their focus when studying AD. By providing different proteins and symptoms that are associated with AD, there are probabilities to allow doctor to diagnose AD at earlier stage and apply more effective medications. We also hope to use our method and result as guidelines for future data collecting. By knowing the filtering process and result details we can enable more efficient data collecting process.

Also, since we did not find our significant SNPs to be significant in our meta analysis result, we wish to look more into how we can solve that problem. We could possibly test our significant SNPs on other data sets besides the three we used to see if we were just unlucky in our test result, or we could also refine the meta analysis data set to be more like our data set so that we can get more similar result.

There are a list of other diseases that we suspect might be associated with LOAD such as Parkinson's disease or Huntington's disease. We can obtain a list of associative genes related to those diseases and find their association with our data set to determine their association with LOAD. A relatively high association with LOAD might suggest potential link between LOAD and those other diseases and that's where we can dig further into.

6 Supplementary Material

References

[1] Ng00073 - genome-wide summary statistics for cognitively defined late-onset alzheimer's disease subgroups, 2018.

[2] The Alzheimer's Association. 2016 alzheimer's disease facts and figures. 2016.

[3] Arief Gusnanto Frank Dudbridge. Estimation of significance thresholds for genomewide association scans. 2008.

[4] National Institute and Aging. Alzheimer's disease genetics fact sheet. 2019.

[5] Ahmet Turan Isik. Late onset alzheimer's disease in older people. 2010.

[6] Amanda J. Myers Jennifer A. Webster, J. Raphael Gibbs. Genetic control of human brain transcript expression in alzheimer disease. 2009.

[7] X Wang R L Minster M M Carrasquillo V S Pankratz S G Younkin A J Saykin G Jun C Baldwin M W Logue J Buros L Farrer M A Pericak Vance

J L Haines R A Sweet M Ganguli E Feingold S T DeKosky O L Lopez M M Barmada M I Kamboh, F Y Demirci. Genome-wide association study of alzheimer's disease. 2012.

[8] Harris S.E. Zhang Q. et al. Marioni, R.E. Gwas on family history of alzheimer's disease. 2018.

[9] math et al. Gwas in plink. https://www.youtube.com/watch?v=7QMSZx3io-Q, Feb 2019.

[10] The International Genomics of Alzheimer's Project. Igap summary statistics, 2013.

[11] Weizmann Institute of Science. Cacng2 gene.

[12] Weizmann Institute of Science. Csnk1e gene.

[13] Weizmann Institute of Science. Lingo2 gene.

[14] Anne Corbett PhD Carol Brayne MD Dag Aarsland MD Emma Jones PhD Prof Clive Ballard MRCPsych, Serge Gauthier MD. Alzheimer's disease. 2011.

[15] Thomas R. Caulfield Chia-Chen Liu Guojun Bu Yu Yamazaki, Na Zhao. Apolipoprotein e and alzheimer disease: pathobiology and targeting strategies. 2019.

	MarkerName	Allele1	Allele2	Zscore	P-value
4527792	rs283811	a	g	12.787	1.940000e-37
2282441	rs111789331	a	t	10.934	7.890000e-28
265147	rs10119	a	g	10.801	3.418000e-27
6878878	rs66626994	a	g	10.355	3.979000e-25
3280941	rs111958034	a	g	-9.933	3.001000e-23
7780692	rs6733839	t	c	9.752	1.803000e-22
5801839	rs75627662	t	c	9.533	1.529000e-21
2084459	rs61737012	a	g	9.109	8.335000e-20
12026484	rs34095326	a	g	8.853	8.538000e-19
8078186	rs471470	a	c	8.484	2.185000e-17
9773150	rs7941541	a	g	8.462	2.619000e-17
8536955	rs78945682	a	g	-8.444	3.079000e-17
4344831	rs7561528	a	g	8.437	3.263000e-17
7810991	rs573167	a	g	8.338	7.585000e-17
9506606	rs968050	t	c	8.096	5.667000e-16
7835873	rs2388334	a	g	8.077	6.629000e-16
2388785	rs9372734	t	c	8.077	6.636000e-16
1366625	rs12202969	a	g	8.066	7.245000e-16
9479997	rs12206087	a	g	8.043	8.789000e-16
7878414	rs9375195	a	g	7.934	2.116000e-15
8172873	rs9375188	t	c	7.926	2.257000e-15
3974213	rs1487441	a	g	7.893	2.949000e-15
8428342	rs9401593	a	c	7.887	3.090000e-15
5790772	rs35114168	a	g	7.883	3.188000e-15
8058813	rs1487445	t	c	7.877	3.343000e-15
13017187	rs1906252	a	c	7.873	3.457000e-15
6005604	rs12204181	t	c	7.845	4.339000e-15
12365616	rs9375225	t	g	7.843	4.392000e-15

Figure 6: Overall Meta-Analysis Result

:

	MarkerName	Allele1	Allele2	P-value
226568	rs1393404	a	g	0.627500
615727	rs1155331	t	c	0.077790
2013534	rs4420638	a	g	0.000680
3242755	rs7043927	a	g	0.467900
9105518	rs429358	t	c	0.000221
10550533	rs9309095	t	c	0.316000
11268026	rs4821510	t	c	0.085530
11511496	rs11114028	a	g	0.017180
11844708	rs867198	t	c	0.103600

Figure 7: Our Significant SNPs in Meta Analysis Result

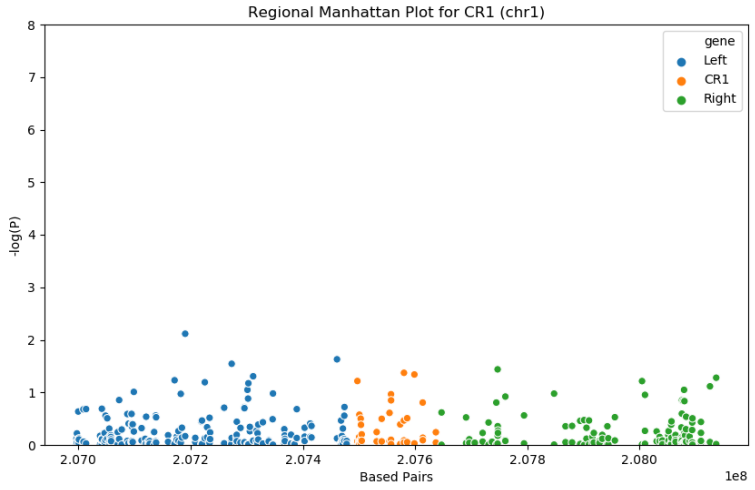

Figure 8: CR1 at Chromosome 1

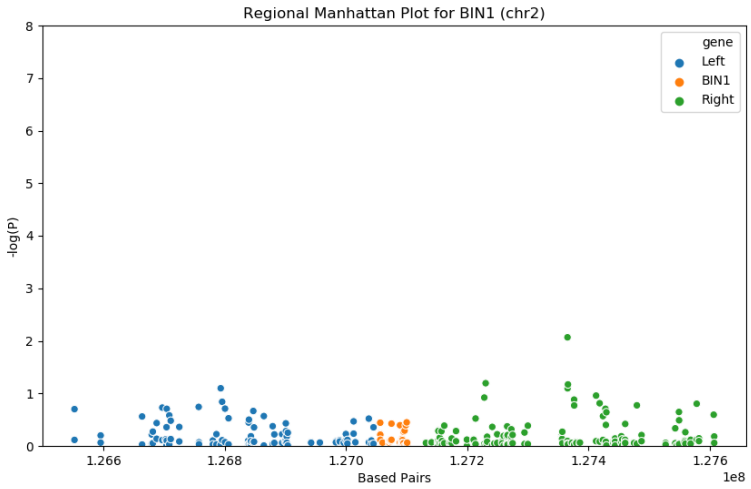

Figure 9: BIN1 at Chromosome 2

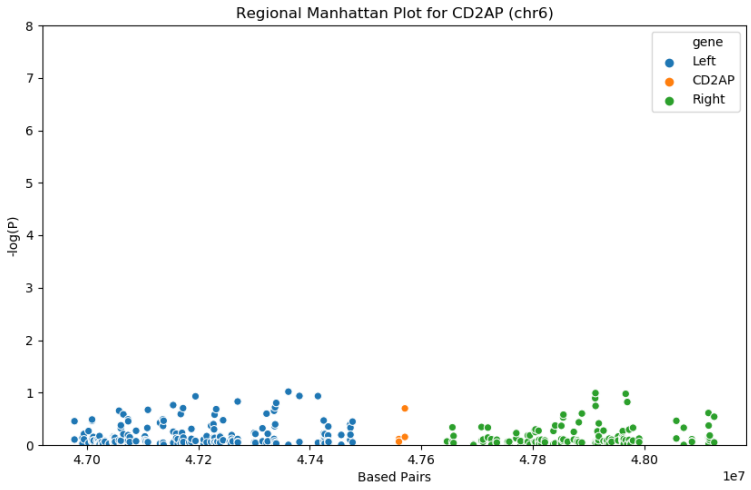

Figure 10: CD2AP at Chromosome 6

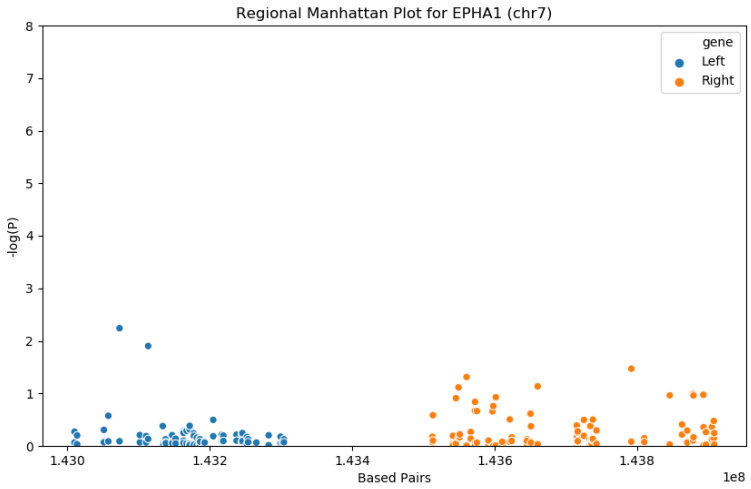

Figure 11: EPHA1 at Chromosome 7

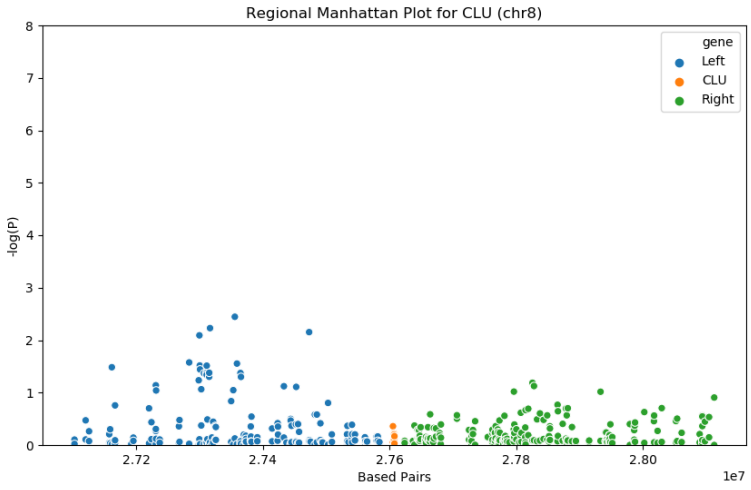

Figure 12: CLU at Chromosome 8

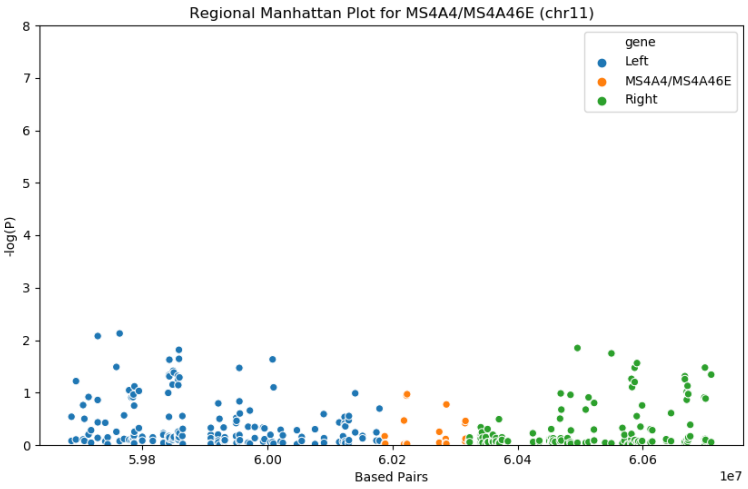

Figure 13: MS4A4/MS4A46E at Chromosome 11

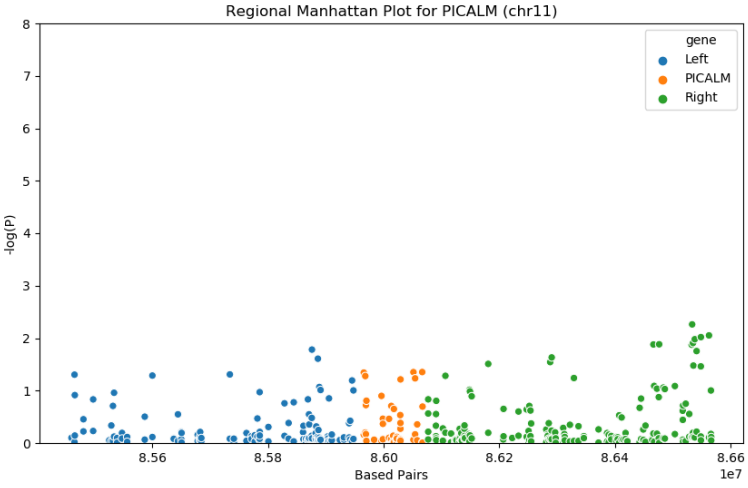

Figure 14: PICALM at Chromosome 11

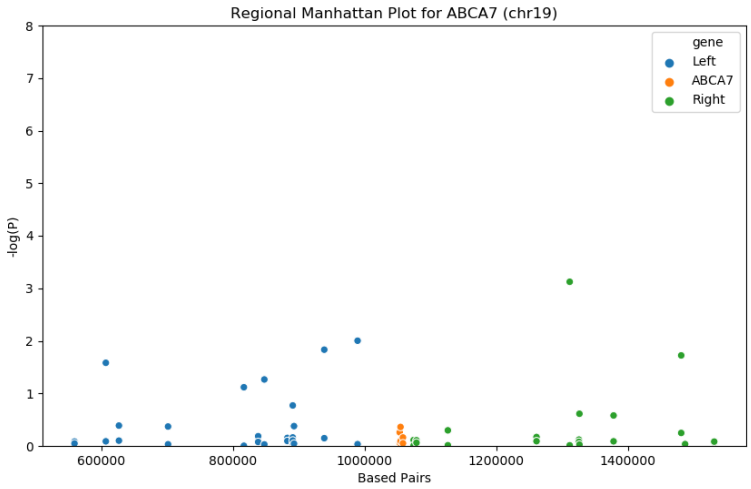

Figure 15: ABCA7 at Chromosome 19

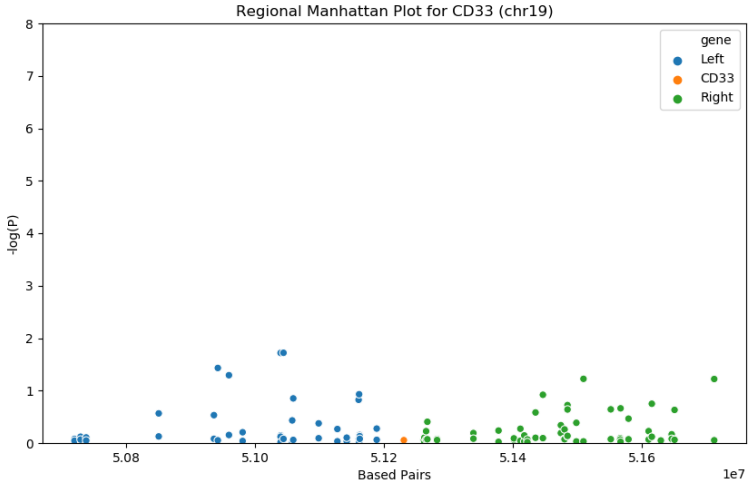

Figure 16: CD33 at Chromosome 19