Discovering Precision AD Biomarkers with Varying Prognosis Effects in Genetics Driven Subpopulations

Brian N. Lee, BA[1], Junwen Wang, PhD[2], Kwangsik Nho, PhD[3], Andrew J. Saykin, PhD[3], Li Shen, PhD[1]
[1]University of Pennsylvania, Philadelphia, PA, USA; [2]Mayo Clinic, Phoenix, AZ, USA; [3]Indiana University, Indianapolis, IN, USA

Abstract

Alzheimer's Disease (AD) is a highly heritable neurodegenerative disorder characterized by memory impairments. Understanding how genetic factors contribute to AD pathology may inform interventions to slow or prevent the progression of AD. We performed stratified genetic analyses of 1,574 Alzheimer's Disease Neuroimaging Initiative (ADNI) participants to examine associations between levels of quantitative traits (QT's) and future diagnosis. The Chow test was employed to determine if an individual's genetic profile affects identified predictive relationships between QT's and future diagnosis. Our chow test analysis discovered that cognitive and PET-based biomarkers differentially predicted future diagnosis when stratifying on allelic dosage of AD loci. Post-hoc bootstrapped and association analyses of biomarkers confirmed differential effects, emphasizing the necessity of stratified models to realize individualized AD diagnosis prediction. This novel application of the Chow test allows for the quantification and direct comparison of genetic-based differences. Our findings, as well as the identified QT-future diagnosis relationships, warrant future investigation from a biological context.

Introduction

Alzheimer's disease (AD) is a complex neurodegenerative disease commonly characterized by memory impairments, cognitive problems, and the presence of both tau and Aβ plaques[1]. As the leading cause of dementia, AD is influenced by environmental and genetic factors[2]. Researchers are facing major challenges to developing effective preventative care and therefore have continued examining the role genetic factors may play in AD etiology and pathogenesis.

Since genetic factors play an important role in AD, genome-wide association studies (GWAS) have been employed to find specific loci and genes that may be instrumental in both AD treatments and prognosis. So far, large case-control GWAS have successfully identified numerous loci susceptible for AD[3]. Additionally, imaging, cognitive, and fluid data have been widely studied to identify quantitative biomarkers that can help predict the status and progression of AD; these continuous measures have also been used to successfully locate AD loci[4,5].

Although many biomarker-based models have increased current knowledge of AD's genetic landscape, to date, no data model or biological paradigm fully explains the heritability of AD risk, necessitating novel approaches for studying AD-related fluid biomarkers, imaging measurements, and cognitive data. Although specific genetic loci have been linked with AD pathololgy or decreased cognition, clinicians still do not understand how specific genetic profiles or traits (i.e. being heterozygous versus homozygous for a risk allele) may specifically affect AD pathology. If these fine genetics-driven pathological differences are well-understood, then researchers can develop predictive models for AD that allow for increasingly accurate and personalized prediction, diagnosis, and treatment. It is impossible to fully realize the potential of precision medicine to predict, diagnose, and treat complex diseases like Alzheimer's without such knowledge.

The Chow test is a powerful tool to perform scalable, biologically-informed analyses determining if different genetic profiles affect a global trait such as future disease diagnosis. The Chow test can be specifically used to determine (1) if an independent variable (i.e. an AD-relevant QT) has different impacts on a specific dependent variable (i.e. future AD diagnosis) across different subgroups of a population (i.e. the stratifying variable, differing allelic dosage of specific genetic variants) and (2) approximate the magnitude of the genetics-based diagnosis differences. The Chow test does so by comparing the parameters of three genetics-stratified linear regression models (one for individuals with no copies of a risk factor allele, one copy of a risk factor allele, two copies of a risk factor allele) to both quantitatively detect the presence of a genetics-based difference and approximate its magnitude. To our current knowledge, this is a

Table 1: List of all SNP's from Jansen et al., 2019 used in Chow test genomic analysis grouped by gene name. The first two rows contain chr2 SNP's and all remaining rows contain chr19 SNP's. SNP's not associated with a specific gene have "NA" listed in the 'Gene' column.

SNP rsID's	*Gene*	*SNP rsID's*	*Gene*
rs7561528	*LOC105373605*	rs28367893	*TRAPPC6A*
rs744373	*NA*	rs12462536	*BLOC1S3*
rs1160985, rs157582, rs8106922, rs2075650, rs157580, rs405697	*TOMM40*	rs2965169	*BCL3*
rs405509, rs769449	*APOE*	rs8111069, rs16979600, rs7257610, rs3786505	*CLPTM1*
rs4420638, rs584007	*APOC1*	rs1114832, rs17643262	*PPP1R37*
rs519825, rs387976, rs12610605, rs6859	*NECTIN2*	rs2965101, rs439401, rs2927438, rs10402271, rs1531517	*NA*

novel application of the Chow test to studying genetics-based differences in future AD diagnosis. The Chow test has been utilized to find genetic subtypes in cancers but is underexplored in the study of AD[6,7].

In this study, the Chow test is used to perform a systematic examination of how specific genetic traits modify associations between AD QT's measured at the baseline visit and future diagnosis at the month 24 and month 36 visits (henceforth referred to as trait-diagnosis associations). We hypothesize that there will be a statistically significant difference between the predictive power of imaging, tau-based, CSF-based, and cognitive biomarkers between individuals with different genetic profiles (defined as having different allelic dosages of key genetic variants), which would greatly support efforts to adopt highly personalized models for AD that more fairly and accurately account for genetic-based differences.

Materials and Methods

Genotyping Reference Data

Since we are specifically interested in examining the differential effects of SNP's known to have a large role in AD pathophysiology and thus diagnosis, we started our analysis by extracting the 30 most significant AD SNP's evaluated in Jansen et al.[8] as determined by the meta-GWAS P value. Jansen et al. performed a large genome-wide association study of clinically diagnosed AD and AD-by-proxy (defined as having a parent with an AD diagnosis, a measure with a high genetic correlation with AD with $r_g = 0.81$) involving 71,770 AD cases and 383,378 healthy controls[8]. This meta-analysis identified 29 novel risk loci, implicating 215 potential causative genes strongly expressed in immune-related tissues and cell types.

The specific SNP's analyzed in this analysis are delineated below in Table 1. A large number of SNP's, including those with extremely low P values, are also associated with known AD risk genes (i.e. APOE, APOC1, and TOMM40) and have been found to be highly associated with dementia-related disorders (including AD) in other studies[9–11].

Genome-wide genotyping data used in this study is sourced from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database[12,13]. ADNI was launched in 2003 as a public-private partnership led by Principal Investigator Michael W. Weiner, MD to test whether serial MRI, PET, and biological markers can be combined with clinical and neuropsychological assessments to accurately measure the progression of mild cognitive impairment (MCI) and early AD[12,13]. Participants were limited to individuals who were subjects of the ADNI cohort. To reduce the likelihood of population stratification effects, only non-Hispanic White participants were involved. As such, there were 1,576 individuals whose genomic sequencing data were included. Demographic data about the individuals included in our analyses can be found in Table 2.

Genotyping data were quality-controlled, imputed and combined[14,15]. Briefly, genotyping was performed on all ADNI participants following the manufacturer's protocol using blood genomic DNA samples and Illumina GWAS arrays (610-Quad, OmniExpress, or HumanOmni2.5-4v1)[16]. Quality control was performed in PLINK v1.90[17] using the following criteria: 1) call rate per marker $\geq 95\%$, 2) minor allele frequency (MAF) $\geq 5\%$, 3) Hardy Weinberg Equilibrium (HWE) test P $\leq$ 1.0E-6, and 4) call rate per participant $\geq 95\%$. Our quality control procedures processed the genetic markers available on the ADNI1 610-Quad panel, where a total of 5,574,300 SNPs were included. From this imputed genotyping data set, we extracted data for the 30 SNP's of interest listed above.

Table 2: ADNI QT-PAD Participant Characteristics. Gender, age (in years), and education (in years) at the baseline are shown. Individuals have been sorted into strata depending on their diagnosis at the baseline visit: healthy control (HC), mild cognitive impairment (MCI), Alzheimer's Disease (AD), or not applicable for individuals who do not have baseline diagnosis data enclosed (N/A). The mean and standard deviation (std) are provided for age and education.

Diagnosis	*HC*	*MCI*	*AD*	*N/A*
Number	461	797	312	6
Gender (% Female)	49.67	40.15	42.95	66.67
Age (Mean±std)	74.47±5.68	73.09±7.54	75.18±7.79	72.15±8.29
Education (Mean±std)	16.42±2.65	16.00±2.82	15.24±2.96	16.75±1.50

Quantiative Trait and Diagnosis Data

Demographic and clinical data used in this analysis was also obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) 1, GO, and 2 cohorts[12]. QT's included in analysis included five cognitive tests (ADAS13, CDRSB, RAVLT.learning, MMSE, FAQ), two PET-based biomarkers (FDG PET, Amyloid PET/AV45), three CSF-based biomarkers (CSF ABETA, CSF TAU, CSF PTAU), and six FreeSurfer neuroimaging features (FS Entorhinal, FS Ventricles, FS MidTemp, FS Fusiform, FS Hippocampus).[18]

The Chow Test Method

The Chow test initially determined if the relationship between a predictor and outcome changed after a major historical event or because of a categorical stratifying variable[19]. More generally, this test determines if the coefficients of two analogous linear regression models built from different strata of a population are equal. If the coefficients significantly differ (as determined by a F test), one can conclude the change is due to the stratifying factor or significant temporal event.

Within the context of this study, we aim to use the Chow test to determine if (1) the ability of specific AD biomarkers measured at baseline levels to predict future diagnoses (at the two-year and three-year visits) varies with the dosage of specific AD SNP's and (2) if the ability of certain imaging phenotypes and biomarkers to predict cognitive outcomes varies with the dosage of specific AD SNP's. As such, we will stratify our population into three separate groups per each of the 30 SNP's depending on an individual's genetic profile (i.e. depending on their individual 'dosage' of the SNP which is in the range $[0, 2]$).

The first set of models evaluated by the Chow test attempted to measure the effect of a specific quantitative biomarker (denoted x_{QT}) on future diagnosis y_{DX} while also controlling for age (c_{AGE}), gender (c_{GENDER}), and education ($c_{EDUCATION}$).

$$\mathrm{y}_{DX} \sim x_{QT} + c_{AGE} + c_{GENDER} + c_{EDUCATION}$$

These models are fitted for each of 16 available biomarkers and all 30 chosen SNP's at two time points (i.e. predicting both two-year and three-year diagnosis). The models are fitted for each of four subpopulations per SNP s: all individuals with allelic dosage $s \in [0, 2]$ (termed the *summative* model), individuals with no doses of the SNP ($s = 0$), individuals with one dosage of the SNP ($s = 1$), and individuals with two copies of the SNP ($s = 2$). Significance between the individual stratified models and the summative model is determined using the F statistic and was expressed as a P value. Significant relationships were chosen using a Bonferroni correction (Chow Test $P < 1.04 \times 10^{-4}$).

Post-Hoc Bootstrapping Analyses

To confirm the differential effects on specific AD biomarkers of certain top SNP's, we bootstrapped ($n = 599$, taking cuts of 80% of the data each time) the regression coefficients calculated for each of the ten statistically significant relationships. Two biologically distinct SNP's were studied: rs7561528 (LOC105373605-chr2), which is associated with hippocampal atrophy and variations in cortical thickness[20], and rs4420638 (APOC1-chr19), which is highly associated with LDL concentrations[21] and is in high linkage disequilibrium (LD) with established APOE SNP rs429358

Table 3: ADNI 3 Participant Characteristics. Gender, age (in years), and education (in years) at the baseline ADNI 3 visit are shown. Individuals have been sorted into strata depending on their diagnosis at the baseline visit: healthy control (HC), mild cognitive impairment (MCI), Alzheimer's Disease (AD), or not applicable for individuals who do not have baseline diagnosis enclosed (N/A). The mean and standard deviation (std) are provided for age and education.

Diagnosis	*HC*	*MCI*	*AD*	*N/A*
Sample Size	221	66	19	0
Gender (% Female)	60.18	36.36	36.84	*N/A*
Age (Mean±std)	70.85±6.04	71.75±7.99	73.79±10.75	*N/A*
Education (Mean±std)	17.00±2.15	16.11±2.42	16.16±2.75	*N/A*

($r^2 = 0.72, D' = 0.96$)[22]. SNP QT's assessed for both SNP's include CDRSB, FAQ, AV45, and FDG because of their noted significance with both of these SNP's in the discovery Chow tests. The regression formula used is identical to those used in the discovery Chow tests.

Post-Hoc Comparison Association Analyses

To verify the utility of stratifying AD participants by allelic dosage of key AD SNP's and calculating three models, we performed two sets of association analyses and compared the resulting P values.

The first set of association analyses will measure the association between a QT and long-term (i.e. month 36) diagnosis using years of education, patient gender, and age as covariates, using data from all patients. The second set of association analyses will stratify the ADNI participants on the allelic dosage of SNPs rs7561528 and rs4420638 before calculating the same associations. After the calculation of all models, the P values of the summative model will be compared to the P values corresponding to the stratified allelic dosage models.

Verification in Independent Cohort

Biological replication of the Chow test results was performed in an independent cohort (sourced from the ADNI 3 cohort) to verify the results of the discovery Chow tests. Clinical (i.e. month 24 diagnosis), demographic (age, gender, education), and biomarker data were downlaoded from ADNI[12]. Genotyping data was processed in a similar fashion to the ADNI 1/GO/2 cohort genotyping data. 29 of the 30 SNP's analyzed in the discovery Chow tests were present in the ADNI 3 genotyping data and were therefore used as stratifying variables; only SNP rs1531517 (chr19-no gene association) was missing. Patient characteristics for the ADNI 3 participants are summarized in Table 3. The predictive relationships assessed are analogous to the relationships analyzed in previous analyses: available QT's (FDG, AV45, ADAS13, MMSE, CDRSB, FAQ, RAVLT.learning) predicted month 24 diagnosis when factoring for patient age, gender, and years of education. Significance was determined using a Bonferroni-corrected P-value threshold $P < 2.46 \times 10^{-4}$.

Results

Our Chow test results are summarized and shown in Figure 1. Several FreeSurfer imaging biomarkers and cognitive scores predicted month 24 and 36 diagnoses with differing regression coefficients between individuals with differing genetic profiles. All statistically significant pairs of outcomes (as determined by a Chow Test P value smaller than the Bonferroni threshold of 1.04×10^{-4}) are marked with a red 'X'.

Strong genetics-based differences involved the cognitive and PET-based QTs predicting month 24 diagnosis ($P = 7.81 \times 10^{-18}$ for the FDG-based prediction, $P = 5.30 \times 10^{-12}$ for the FAQ-based prediction) and month 36 diagnosis (FDG: $P = 2.59 \times 10^{-11}$, FAQ: $P = 2.67 \times 10^{-21}$) when stratified by the dosage of chr19 SNP rs4420638. Differences in the dosage of chromosome 2 SNP rs7561528 also affected predictive relationships involving cognitive and PET-based QT's: CDRSB-based predictions of month 24 and month 36 diagnosis had Chow test P values of 2.51×10^{-19} and 4.96×10^{-19}, respectively, and AV45-based relationships were also significant ($P = 6.55 \times 10^{-5}$ and 6.08×10^{-5} for the month 24 and 36 predictions, respectively). The combination of these results confirms the hypothesized presence of significant genetics-based differences, particularly when creating models to predict baseline

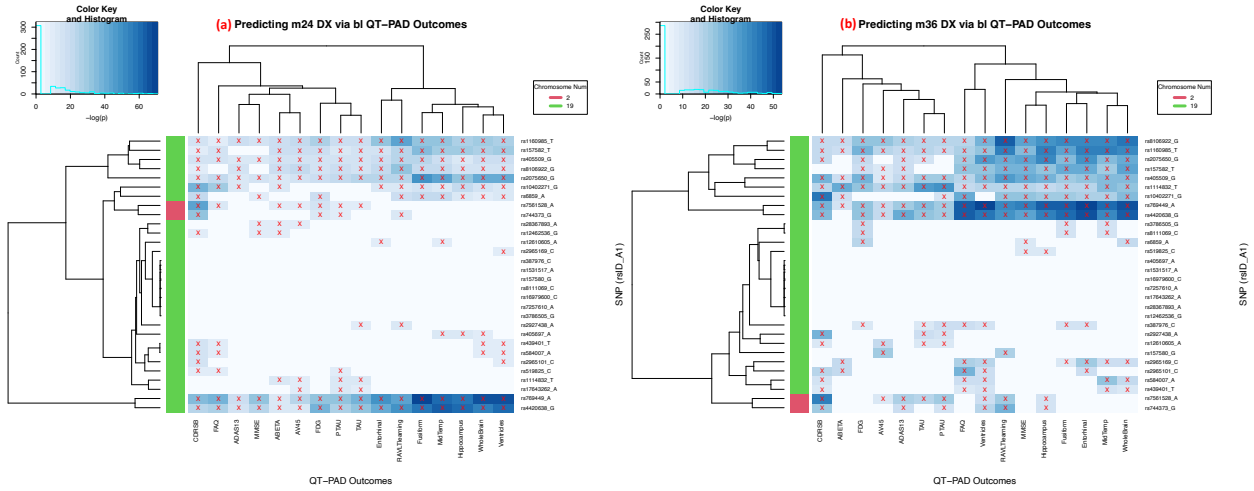

Figure 1: Heat maps showing results of Chow test. Regression: predict month 24 diagnosis (subfigure a) or month 36 diagnosis (subfigure b) using a variety of baseline imaging, CSF, and cognitive biomarkers (horizontal axis) when stratifying on a variety of key AD genetic markers (vertical axis). Covariates included age, years of education, and patient gender as covariates. Vertical color bar represents chromosome associated with a specific SNP (green for chr19, red for chr2) and shade of cells denotes relative -log(Chow Test P). Significance was determined by a Bonferroni threshold ($P < 1.04 \times 10^{-4}$) with significant relationships denoted as X.

and future cognitive status using QT's.

Bootstrapping ($n = 599$) confirmed the significance of genetic-based differences in QT-diagnosis predictions[23]. To highlight a range of predictors and genetic markers, the relationships using AV45, FDG, CDRSB, and FAQ to predict month 24 and 36 diagnosis when stratifying upon the allelic dosage of SNP's rs7561528 and rs4420638 were bootstrapped, with the bootstrapped regression coefficients shown in Figures 2.

Bootstrapping reveals noticeable differences in regression coefficients. Differences are especially stark for the AV45-based, FDG-based, CDRSB-based, and FAQ-based predictions of month 24 diagnosis when stratifying on rs4420638 (Chow test $P = 4.95 \times 10^{-6}$, $P = 7.81 \times 10^{-18}$, $P = 2.78 \times 10^{-11}$, and $P = 5.30 \times 10^{-12}$, respectively). These differences were also notable for the analogous month 36 diagnosis predictions (Chow test $P = 7.68 \times 10^{-8}$, $P = 2.59 \times 10^{-11}$, $P = 3.55 \times 10^{-12}$, and $P = 2.67 \times 10^{-21}$ for AV45, FDG, CDRSB, and FAQ, respectively). Notably, the directionality of the regression coefficients when increasing in allelic dosage is also consistent with rs4420638's status as a risk factor allele: it is expected for amyloid-beta concentrations as measured by AV45 to rise and cognitive test scores to fall with an increased allelic dosage[24]. Less stark but still-significant differences were also notable in the rs7561528 bootstrapped analyses, with the stronger differences noted in the month 36 predictive relationships (AV45-based prediction P value of $P = 6.08 \times 10^{-5}$ and CDRSB-based prediction P value of $P = 4.96 \times 10^{-19}$). The directionality of the regression coefficients as allelic dosage increases from 0 to 2, particularly in the month 36 diagnoses, is also consistent with current knowledge indicating the SNP is a protective factor for AD[25].

To visually compare the differences noted in the Chow tests and bootstrapped regression coefficients, a series of scatterplots were made (Figure 3). These plots conveniently depict a genetics-difference-caused change in the regression coefficients as differences in the slope of the three allelic-dosage-based linear regression models' plotted lines.

Individual points (individuals with zero copies of the SNP's affect allele in orange, individuals with one copy of the affect allele in blue, individuals with two copies of the affect allele in green) represent data from the ADNI cohorts; each participant's QT is plotted against an adjusted diagnosis value. The adjusted diagnosis is calculated by subtracting the intercept from the QT-future diagnosis regression model (using all participants, as opposed to using three different intercepts from the participants with 0, 1, or 2 copies of the specified SNP effect allele) from an encoded diagnosis score (HC was encoded as 0, MCI diagnosis was encoded as 1, and AD diagnosis was encoded as 2). Plotting the adjusted diagnosis allows for a more informative visualization – slope differences between the stratified versus additive

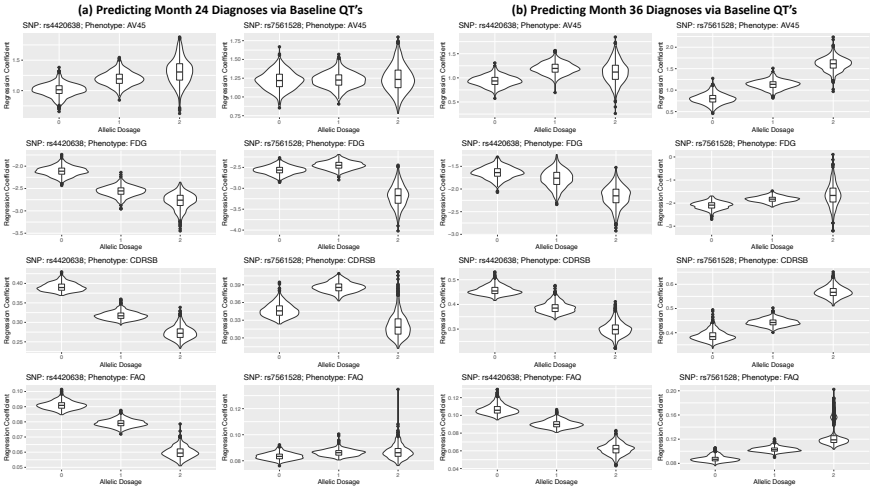

Figure 2: Violin plots showing the results of bootstrapping analysis ($n = 599$; randomly sampling 80% of the available data per diagnosis time point, QT, and SNP per iteration) to evaluate the results of the Chow tests. A select number of correlations from the Chow tests (see Figure 1) were chosen. Subfigure (a) shows predictions of month 24 diagnosis and subfigure (b) shows predictions of month 36 diagnosis. The box plot shows the median and IQR of calculated regression coefficients and specific points highlight outliers.

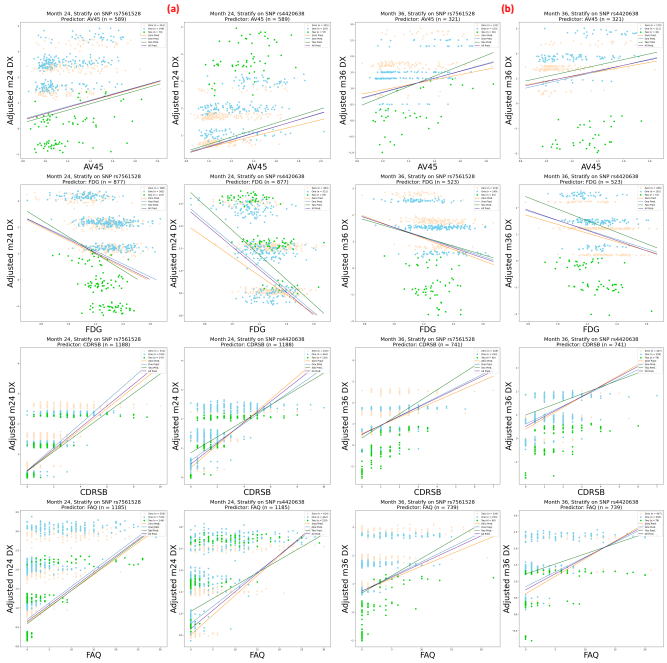

Figure 3: Example linear predictive models of future diagnosis (subfigure (a) corresponds to month 24 diagnosis predictions and subfigure (b) corresponds to month 36 diagnosis predictions) learned from subjects with varying allelic dosages of SNP's rs4420638 and rs7561528 and all subjects using baseline data. The vertical axis represents an adjusted diagnosis encoding after regressing out the effects of age, gender, and years of education. The horizontal axis represents the respective QT predictor.

models are significantly more apparent. Then, regression lines for the QT-diagnosis outcome were plotted using the data from all participants (in purple), participants with zero copies of the SNP affect allele (in orange), participants

with one copy of the affect allele (in blue), and participants with two copies of the affect allele (in green).

In the scatterplots shown in Figure 3, genetics-related differences can be seen via the different slopes of the summative versus three stratified regression lines. The differences are especially noticeable in the month 36 AV45 plots (the first column of subfigure b) and the month 24 and 36 CDRSB plots (the entire third column) but less so in the FDG and FAQ plots. Additionally, the stratified-genetics regression lines better represent trends in their respective strata than the summative regression line does for specific genetic subpopulations (i.e. individuals with a specified allelic dosage of one of the two key SNP's). The presence of these genetics-based differences corroborate the significant P value produced via the Chow test. As such, one can infer that there likely exists a significant biological effect of specific genotypes on the predictive relationships between AD-related QT's and future diagnosis. Knowledge of these differences can be integrated to create more refined, individualized genetic models for AD diagnosis predictions.

To establish the utility of using multiple stratified genetics-based models for long-term (i.e. month 36) AD diagnosis, four sets of association analyses (using participants with zero copies of the effect allele, one copy of the effect allele, two copies of the affect allele, and all participants) predicting an encoded diagnosis score using a QT were performed. Figure 4 displays the results (i.e. P-values) of these analyses side-by-side for direct comparison. Data from both investigated SNP's (rs4420638 in subfigure a, rs7561528 in subfigure b) were used. As expected, the large majority of trait-outcome associations were deemed statistically significant ($P < 3.13 \times 10^{-3}$) when using all patient data. However, doing so glosses over differences between the different groups of individuals with varying genetic profiles. For example, with the AV45-based prediction of month 36 diagnosis, the association is statistically significant in only participants without or heterozygous for the rs4420638 affect allele. Simply using a standard, Mendelian-based additive understanding of this SNP's effect on long-term diagnosis may lead to incorrect predictions that would have been realized with a series of stratified models. A similar scenario exists with the FDG-based prediction of month 36 diagnosis; comparative analyses show the highly-significant association between trait and diagnosis is only significant in individuals with zero or one copy of the effect allele, emphasizing the utility of stratification-based analyses like the Chow test.

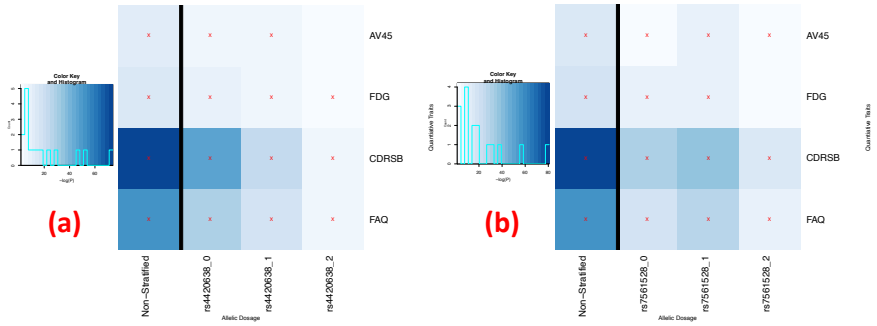

Figure 4: Heat maps showing results of confirmatory association tests. Regression: predict month 36 diagnosis using a variety of imaging and cognitive biomarkers (vertical axis) when factoring for age, years of education, and patient gender as covariates. Horizontal color bar represents population used (all participants, participants with zero copies of affect allele, one copy of affect allele, or two copies of effect allele) and shade of cells denotes relative -log(Association Test P). Significance was determined by a Bonferroni threshold ($P < 3.13 \times 10^{-3}$) with significant relationships denoted as X.

Last but not least, the Chow test findings reported here were replicated in an independent cohort. A large number of associations present in the initial month 24 and month 36 Chow test analyses were reassuringly significant in this replication, validating the initial findings. Notably, FDG-based ($P = 1.53 \times 10^{-10}$), AV45-based ($P = 1.16 \times 10^{-15}$), CDRSB-based ($P = 3.79 \times 10^{-6}$), MMSE-based ($P = 2.47^{-6}$), and RAVLT-based ($P = 6.32 \times 10^{-11}$) predictions stratifying on the allelic dosage of rs4420638 were significant in both the replication and the initial analyses. The FDG-based prediction of month 24 diagnosis stratifying on rs7561528 was also significant in both the replication ($P = 1.91 \times 10^{-10}$), duly highlighting the strength of the genetics-based differences found by the Chow test.

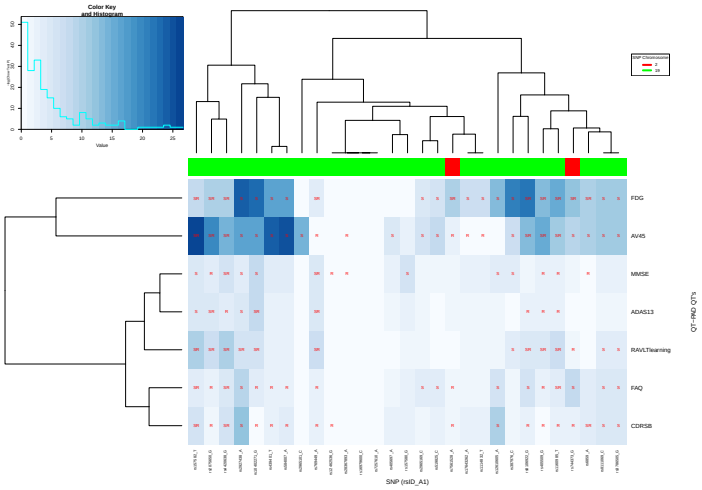

Figure 5: Heat map showing results of biological replication Chow test using ADNI 3 data. Regression: predict month 24 diagnosis using a variety of imaging and cognitive biomarkers (vertical axis) when stratifying on 29 of 30 initial AD genetic markers (horizontal axis). Covariates included age, years of education, and patient gender. Horizontal color bar represents chromosome associated with specific SNP's (green for chr19, red for chr2) and shade of cells denotes relative -log(Chow Test P). Significance was determined by a Bonferoni-corrected P-value threshold ($P < 2.46 \times 10^{-4}$); significant relationships are denoted with a red 'S', relationships significant in the discovery Chow tests are denoted with a red 'R' (see Figure 1), and relationships significant in both the replication and initial Chow test analyses are denoted with a red 'SR'.

Discussion

The Chow test has been successfully used to identify QT-diagnosis relationships with statistically significant differences between individuals with varying genetic profiles. The most striking disparities occur in relationships involving cognitive and imaging-measured traits. There were also notable genetics-based differences in a handful of neuroimaging features. The magnitude of the genetics-based differences has been confirmed and visualized through a variety of visualizations (violin plots, scatterplots) and post-hoc comparative analyses (bootstrapping and genetics-stratified association tests). Additionally, given that many of the quantitative cognitive, imaging, and fluid biomarkers are closely tied to AD pathology, it is reasonable to expect significant genetics-based differences, corroborating the abilities of a stratified approach to more accurately model indicators of AD pathology.

To the best of our knowledge, finding and quantifying genetics-based differences as performed in this manuscript is a novel application of the Chow test. The quantification of genetic differences in imaging traits and cognitive test scores can improve current models of AD, allowing for more precise diagnosis predictions. Using a larger, more diverse, and more balanced cohort, future analyses can perform more targeted genetic studies of AD across the entire genome, ushering an era of precision medicine and increasingly specialized diagnosis and treatment.

The Chow test also has the unique methodological advantage of quantifying the magnitude of noted genetics-based differences. This is a useful measure not ordinarily attainable via standard stratified analyses (see Figure 4). Although it is possible to notice significant differences via direct comparison of P values from different strata, as is done in Figure 4, the Chow test's utilization of the F statistic can facilitate the comparison of different relationships (i.e. comparing individual QT-diagnosis pairs across different time points and when stratifying on different SNP's).

A majority of the QT-future diagnosis associations with statistically significant Chow test P values specifically involve the CDRSB, ADAS13, MMSE, and FAQ cognitive outcomes. This is likely due to the increased granularity of these two clinical scales as well as the larger sample sizes of these scales, given their prevalence in the clinic. Given their ties to clinical measures of cognitive impairment, these scores (as well as other clinical/pathophysiological scales such as the MOCA or Braak stage score) may also serve as ideal proxies for AD diagnosis in the future and allow researchers to investigate the changing ability of the non-cognitive biomarkers to directly predict AD diagnosis.

A large number of genetic differences also utilized imaging-based biomarkers (including both the AV45 and FDG PET-based measurements and volumetric FreeSurfer MRI-based phenotypes). These MRI and PET-based measurements are a common way to track AD pathology over time and, similar to many of the clinical batteries, can provide a level of extra granularity that can be helpful in quantitative analyses. Because Alzheimer's disease is a neurodegenerative disorder, these quantitative measurements can be especially effective methods to track long-term changes in specific brain regions and amyloid burden, allowing for more targeted downstream analyses.

This study differs from other stratified studies due to the use of the F statistic to evaluate the significance of the difference between the effect sizes of a pair of models. In doing so, our analysis corroborates previous findings highlighting the anatomical and corresponding pathophysiological differences between brains of individuals with differing genetic profiles. Perhaps the differences noted in our analyses here could inform the models used to predict AD pathogenesis and reinforce the utility of stratified analyses such as the Chow test.

Indeed, varying P values between subjects of differing genetic profiles highlight the usefulness of stratified analyses; only a stratified analysis would accurately reflect genetic differences in AD pathophysiology that must be accounted for in predictive models and patient care. If sufficient data allow, researchers should use independent models for patients with different genetic profiles when creating predictive models for key AD biomarkers associated with AD diagnosis or pathology. This is because combining all subjects in a single analysis – the standard approach in large genome-wide association studies and machine learning-based approaches – may overemphasize associations in participants homozygous or heterozygous for an effect allele. Only in stratified analyses are these differential associations visible.

There may be some concerns regarding the sample size used in our study as well as the balance of individuals with different diagnosis codes, particularly in the verification study. Ideally, it would be possible to utilize a full data set with samples from multiple AD biomarkers with a suitably large cohort with a fairly distributed number of patients at all stages of AD progression. However, doing so is likely difficult, due to the ongoing nature of the ADNI studies and the logistical difficulties involved with longitudinal clinical trials. Future studies would involve efforts to verify the results of these analyses, perhaps in a different cohort of patients entirely or via use of updated statistics and biomarker measurements from the ADNI data.

Conclusions

In conclusion, genetic analysis in conjunction with AD biomarker data via the Chow test identified several SNP's coupled with precision AD biomarkers with varying prognosis effects in the corresponding genotype groups. The largest differential effects were noticed when utilizing cognitive outcomes, such as the MMSE, CDRSB, and FAQ scales, to predict long-term (i.e. month 24 or 36) diagnosis while accounting for the covariates of age, gender, and years of education. These findings warrant additional investigation in order to determine if the differences found may reveal disease heterogeneity or be used to facilitate precision medicine-grounded diagnostic measures and future treatments.

Acknowledgements

This work was supported in part by the National Institutes of Health [R01 LM013463, U01 AG066833, U01 AG068057, R01 AG071470]; and National Science Foundation IIS 1837964. Data used in this study were obtained from the Alzheimer's Disease Neuroimaging Initiative database (adni.loni.usc.edu), which was funded by NIH U01 AG024904.

References

1. Dubois B, Hampel H, Feldman HH, et al. Preclinical Alzheimer's disease: Definition, natural history, and diagnostic criteria. *Alzheimer's dementia : the journal of the Alzheimer's Association*, 12(3):292–323, mar 2016.
2. C. A. Lane, J. Hardy, and J. M. Schott. Alzheimer's disease. *European Journal of Neurology*, 25(1):59–70, jan 2018.
3. Andrew J. Saykin, Li Shen, et al. Genetic studies of quantitative MCI and AD phenotypes in ADNI: Progress, opportunities, and plans. 11(7):792–814, jul 2015.
4. Brian Lee, Xiaohui Yao, Li Shen, and Alzheimer's Disease Neuroimaging Initiative. Genome-Wide association

study of quantitative biomarkers identifies a novel locus for alzheimer's disease at 12p12.1. *BMC Genomics*, 23(1):85, January 2022.

5. Li Shen and Paul M. Thompson. Brain imaging genomics: Integrated analysis and machine learning. *Proceedings of the IEEE*, 108(1):125–162, 2020.
6. Wenhua Liang, Jiaxi He, Yaxing Shen, et al. Impact of Examined Lymph Node Count on Precise Staging and Long-Term Survival of Resected Non–Small-Cell Lung Cancer: A Population Study of the US SEER Database and a Chinese Multi-Institutional Registry. *https://doi.org/10.1200/JCO.2016.67.5140*, 35(11):1162–1170, dec 2016.
7. Jianfei Fu, Hang Ruan, Hongjuan Zheng, et al. Impact of old age on resectable colorectal cancer outcomes. *PeerJ*, 7(2):e6350, feb 2019.
8. Iris E. Jansen, Jeanne E. Savage, Kyoko Watanabe, et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nature Genetics 2019 51:3*, 51(3):404–413, jan 2019.
9. Lambert JC, Ibrahim-Verbaas CA, Harold D, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nature genetics*, 45(12):1452–1458, dec 2013.
10. Schwartzentruber J, Cooper S, Liu JZ, et al. Genome-wide meta-analysis, fine-mapping and integrative prioritization implicate new Alzheimer's disease risk genes. *Nature genetics*, 53(3):392–402, mar 2021.
11. Derrek P. Hibar, Jason L. Stein, et al. Common genetic variants influence human subcortical brain structures. *Nature*, 520(7546):224–229, apr 2015.
12. ADNI — Alzheimer's Disease Neuroimaging Initiative, 2012.
13. Li Shen, Paul M. Thompson, et al. Genetic analysis of quantitative phenotypes in AD and MCI: Imaging, cognition and biomarkers. *Brain Imaging and Behavior*, 8(2):183–207, 2014.
14. Xiaohui Yao, Shan Cong, et al. Regional imaging genetic enrichment analysis. *Bioinformatics*, 36(8):2554–2560, 12 2019.
15. Xiaohui Yao, Shannon L Risacher, et al. Targeted genetic analysis of cerebral blood flow imaging phenotypes implicates the inpp5d gene. *Neurobiology of 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 & dementia*, 6(3):265–273, 2010.
17. Shaun Purcell, Benjamin Neale, et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, 81(3):559–575, 2007.
18. Portland Institute for Computational Science. Alzheimer's Disease Modelling Challenge: Modelling the progression of Alzheimer's disease. `http://www.pi4cs.org/qt-pad-challenge`, 2012.
19. Gregory C. Chow. Tests of Equality Between Sets of Coefficients in Two Linear Regressions. *Econometrica*, 28(3):591, jul 1960.
20. Hema Sree G N S, V Lakshmi Prasanna Marise, Kshreeraja S Satish, et al. Untangling huge literature to disinter genetic underpinnings of alzheimer's disease: A systematic review and meta-analysis. *Ageing Res. Rev.*, 71(101421):101421, November 2021.
21. Cristen J Willer, Serena Sanna, Anne U Jackson, et al. Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nat. Genet.*, 40(2):161–169, February 2008.
22. Fiona Cunningham, James E Allen, Jamie Allen, et al. Ensembl 2022. *Nucleic Acids Research*, 50(D1):D988–D995, 11 2021.
23. R.R. Wilcox. *Fundamentals of Modern Statistical Methods: Substantially Improving Power and Accuracy*. Springer New York, 2010.
24. Hao Li, Sally Wetten, Li Li, et al. Candidate single-nucleotide polymorphisms from a genomewide association study of alzheimer disease. *Arch. Neurol.*, 65(1):45–53, January 2008.
25. Futao Zhou and Dong Haina. The bridging integrator 1 gene rs7561528 polymorphism contributes to alzheimer's disease susceptibility in east asian and caucasian populations. *Clin. Chim. Acta*, 469:13–21, June 2017.