

University of Kentucky

UKnowledge

Theses and Dissertations--Epidemiology and
Biostatistics

College of Public Health


2021

Investigations into the Genetics of Mixed Pathologies in Dementia

Adam Dugan

University of Kentucky, adam.dugan@uky.edu

Author ORCID Identifier:

 <https://orcid.org/0000-0001-5783-3781>

Digital Object Identifier: <https://doi.org/10.13023/etd.2021.315>

[Right click to open a feedback form in a new tab to let us know how this document benefits you.](#)

Recommended Citation

Dugan, Adam, "Investigations into the Genetics of Mixed Pathologies in Dementia" (2021). *Theses and Dissertations--Epidemiology and Biostatistics*. 31.

https://uknowledge.uky.edu/epb_etds/31

This Doctoral Dissertation is brought to you for free and open access by the College of Public Health at UKnowledge. It has been accepted for inclusion in Theses and Dissertations--Epidemiology and Biostatistics by an authorized administrator of UKnowledge. For more information, please contact UKnowledge@lsv.uky.edu.

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Adam Dugan, Student

Dr. David W. Fardo, Major Professor

Dr. Heather M. Bush, Director of Graduate Studies

INVESTIGATIONS INTO THE GENETICS OF MIXED PATHOLOGIES IN
DEMENTIA

DISSERTATION

A dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy in the
College of Public Health
at the University of Kentucky

By

Adam Joseph Dugan

Lexington, Kentucky

Co- Directors: Dr. David W. Fardo, Professor of Biostatistics

and Dr. Olga A. Vsevolozhskaya, Assistant Professor of Biostatistics

Lexington, Kentucky

2021

Copyright © Adam Joseph Dugan 2021
<https://orcid.org/0000-0001-5783-3781>

ABSTRACT OF DISSERTATION

INVESTIGATIONS INTO THE GENETICS OF MIXED PATHOLOGIES IN DEMENTIA

Alzheimer's disease (AD) is an irreversible, progressive brain disorder that leads to a loss of memory and thinking skills. While tremendous progress has been made in our understanding of the genetics underlying AD, currently known genetic variants explain only approximately 30% of the heritable risk of developing AD. One hurdle to AD research is that it can only be definitively diagnosed at autopsy, making cruder, clinic-based diagnoses more common. In recent years, several brain pathologies that mimic AD's clinical presentation have been identified including brain arteriolosclerosis, hippocampal sclerosis (HS), and, most recently, limbic-predominant age-related TDP-43 encephalopathy (LATE). It has become increasingly clear that "pure AD" is rare and mixed pathologies (i.e., having two or more concomitant pathologies) are very prevalent. In this dissertation, I will present two investigations into the genetics of mixed pathologies based on autopsy-confirmed neuropathological phenotypes along with a new statistical method with an application to the genetics of mixed pathologies. The first investigation looks at a recently identified clinical AD risk locus and finds novel associations with two AD mimics and, importantly, no associations with AD-related pathologies, which suggests that the locus may be preferentially associated with non-AD dementia. The second investigation looks at the shared genetics of two related AD mimics, HS and LATE, and replicates earlier findings while also identifying several novel functional variants for both HS and LATE. The new statistical method leverages models from the branch of statistics known as functional data analysis to create a gene-level genetic pleiotropy test statistic. An extensive simulation study found that the test statistic outperforms competing methods in small sample, modest effect size scenarios and when applied to real-world data identified a novel joint association between HS and LATE and the GRN gene. All investigations were able to leverage neuropathologically-confirmed endophenotypes to identify novel genetic associations with several AD mimics, adding to the growing body of literature on the complex genetics underling neurodegenerative disease and the statistical methods available for such studies.

KEYWORDS: Alzheimer's Disease, Hippocampal Sclerosis, Limbic-predominant Age-related TDP-43 Encephalopathy, Mixed Pathologies, Genetic Pleiotropy, Functional Data Analysis.

Adam Joseph Dugan

07/23/2021

Date

INVESTIGATIONS INTO THE GENETICS OF MIXED PATHOLOGIES IN
DEMENTIA

By
Adam Joseph Dugan

David W. Fardo

Co-Director of Dissertation

Olga A. Vsevolozhskaya

Co-Director of Dissertation

Heather M. Bush

Director of Graduate Studies

07/23/2021

Date

TABLE OF CONTENTS

LIST OF TABLES	vi
LIST OF FIGURES	viii
CHAPTER 1. Introduction.....	1
1.1 Alzheimer’s Disease	1
1.2 Other Neurodegenerative Diseases	1
1.3 Genetics.....	2
1.4 Dissertation Outline	3
CHAPTER 2.Association between WWOX/MAF variants and dementia-related neuropathologic endophenotypes.....	5
2.1 Abstract	5
2.2 Introduction.....	5
2.3 Materials and Methods.....	7
2.3.1 Study Participants	7
2.3.2 Neuropathological Endophenotype Definitions.....	8
2.3.3 Quality Control of Genotype Data.....	9
2.3.4 Variant-Level Associations.....	10
2.3.5 Variant Prioritization and Downstream Analyses.....	11
2.3.6 Sensitivity Analyses.....	12
2.4 Results.....	13
2.4.1 Variant-Level Associations.....	13
2.4.2 Variant Prioritization and Downstream Analyses.....	14
2.4.3 Sensitivity Analyses.....	18
2.4.3.1 Varying WWOX/MAF Flanking	18
2.4.3.2 Age of Death 75+.....	18
2.4.3.3 Adjusting for ADNC.....	19
2.5 Discussion.....	20
2.6 Conclusion	22
CHAPTER 3. Analysis of genes (TMEM106B, GRN, ABCC9, KCNMB2, and APOE) implicated in risk for LATE-NC and hippocampal sclerosis provides pathogenetic insights.....	28
3.1 Abstract.....	28

3.2	Introduction.....	29
3.3	Material and Methods	31
3.3.1	Study Participants	31
3.3.2	Neuropathological Endophenotype Definitions.....	32
3.3.3	Quality Control of Genotype Data.....	33
3.3.4	Variant-Level Associations.....	34
3.3.5	Gene-Based Associations.....	35
3.3.6	SNV Prioritization and Follow-Up Analyses	36
3.3.7	Sensitivity Analyses.....	37
3.4	Results.....	37
3.4.1	Gene-Based Associations.....	38
3.4.2	Prioritized SNVs and Follow-Up Analyses	39
3.4.3	SNV-Level Regression Analyses.....	41
3.5	Discussion.....	43
CHAPTER 4. A new functional F-statistic for gene-based inference involving multiple phenotypes.....		61
4.1	Abstract.....	61
4.2	Introduction.....	61
4.3	Methods.....	63
4.3.1	Functional F-Statistics	63
4.3.1.1	The Function-on-Scalar Regression Model	63
4.3.1.2	The Shen and Faraway Functional F-Statistic	64
4.3.1.3	Newly Proposed Functional F-Statistic	64
4.3.1.4	Fitting FoSR Models.....	66
4.3.2	The Gene Association with Multiple Traits Test.....	66
4.3.3	The Aggregated Cauchy Association Test.....	68
4.3.4	Simulations	69
4.4	Results.....	71
4.4.1	Simulations	71
4.4.1.1	Type I Error.....	71
4.4.1.2	Power	72
4.4.2	Application to Neurodegenerative Disease Data	73
4.5	Discussion.....	75
CHAPTER 5. Conclusion.....		86
5.1	Summary.....	86
REFERENCES		89

VITA..... 96

LIST OF TABLES

Table 2.1. Individual characteristics stratified by endophenotype status for National Alzheimer's Coordinating Center (NACC) and Religious Orders Study and Rush Memory and Aging Project (ROSMAP) participants.....	23
Table 2.2. Adjusted results for previously published HS and clinical AD variants in the WWOX/MAF locus. All analyses adjusted for age at death, sex, Alzheimer's Disease Genetics Consortium (ADGC) cohort or Religious Orders Study and Memory and Aging Project (ROSMAP) study, and the first three genetic principal components. The rs55751884 results are reported assuming a recessive mode of inheritance (MOI) as that was the MOI with the strongest association in Nelson et al., 2014. The rs450674 results are reported assuming an additive MOI as that was the MOI reported in Bellenguez et al., 2020.....	24
Table 2.3. Variant-level results for variants with uncorrected meta-analytic p-values that met the Bonferroni-corrected threshold for significance for the WWOX/MAF locus \pm 250kb. All analyses adjusted for age at death, sex, Alzheimer's Disease Genetics Consortium (ADGC) cohort or Religious Orders Study and Memory and Aging Project (ROSMAP) study, and the first three genetic principal components.....	25
Table 3.1. Participant characteristics stratified by hippocampal sclerosis (HS) and limbic-predominant age-related TDP-43 encephalopathy neuropathological changes (LATE-NC) case status.....	47
Table 3.2. Aggregated Cauchy association test (ACAT) gene-based p-values for hippocampal sclerosis (HS) and limbic-predominant age-related TDP-43 encephalopathy neuropathological changes (LATE-NC). Each gene is flanked by 10kb. All SNV-level analyses were adjusted for sex, age at death, cohort/study, and the first three genetic principal components and meta-analyzed across National Alzheimer's Coordinating Center (NACC) and Religious Orders Study and Rush Memory and Aging Project (ROSMAP) participants.....	48
Table 3.3. Most significant expression quantitative trait loci (eQTL) p-values for ABCC9 in BRAINEAC and GTEx databases.....	49
Table 3.4. Adjusted effects of single nucleotide variants (SNV) on limbic-predominant age-related TDP-43 encephalopathy neuropathological change (LATE-NC). All models adjust for sex, age at death, first three principal components and cohort/study. For rs1990622, rs7781670, and rs704178, the effect alleles are the risk-associated alleles and not the minor alleles.....	50
Table 3.5. Adjusted effects of single nucleotide variants (SNV) on hippocampal sclerosis (HS). A separate regression model was fit for each variant, mode of inheritance (MOI), and limbic-predominant age-related TDP-43 encephalopathy neuropathological change (LATE-NC) adjustment. All models also adjust for sex, age at death, first three principal	

components and cohort/study. For rs1990622, rs7781670, and rs704178, the effect alleles are the risk-associated alleles and not the minor alleles.....51

Table 4.1. Type 1 error at $\alpha = 0.05$ for each method stratified by gene size, number of observations, correlation among phenotypes, and correlation disturbance.....79

Table 4.2. Gene-based results from the applied analysis looking at the pleiotropic effects of several genes on HS and LATE. P-values < 0.05 are in bold.....80

LIST OF FIGURES

Figure 2.1. LocusZoom plots of the WWOX/MAF region \pm 250kb for A neuritic plaques and B neurofibrillary tangles, both assuming an additive MOI. Meta-analytic variant-level p-values were adjusted for age at death, sex, Alzheimer's Disease Genetics Consortium (ADGC) cohort or Religious Orders Study and Memory and Aging Project (ROSMAP) study, and first three genetic principal components and meta-analyzed across the NACC and ROSMAP cohorts. The horizontal line at 4.44 represents the Bonferroni-corrected threshold for significance for the WWOX/MAF locus \pm 250kb. The blue region on the gene window highlights the location of rs55751884, the variant previously found to be genome-wide suggestive for HS; the green region on the gene window highlights the location of rs62039712, the variant previously found to be genome-wide significant for clinical AD; and the red region on the gene window highlights the location of rs450674, an MAF variant recently found to be associated with clinical AD.....26

Figure 2.2. LocusZoom plots of the WWOX/MAF region \pm 250kb for A hippocampal sclerosis (HS) assuming a recessive mode of inheritance (MOI), B limbic-predominant age-related TDP-43 encephalopathy neuropathological changes (LATE-NC) assuming an additive MOI, and C brain arteriolosclerosis assuming a dominant MOI. Variant-level p-values were adjusted for age at death, sex, Alzheimer's Disease Genetics Consortium (ADGC) cohort or Religious Orders Study and Memory and Aging Project (ROSMAP) study, and first three genetic principal components. The horizontal line at 4.44 represents the Bonferroni-corrected threshold for significance for the WWOX/MAF locus \pm 250kb. The blue region on the gene window highlights the location of rs55751884, the variant previously found to be genome-wide suggestive for HS; the green region on the gene window highlights the location of rs62039712, the variant previously found to be genome-wide significant for clinical AD; and the red region on the gene window highlights the location of rs450674, an MAF variant recently found to be associated with clinical AD.....27

Figure 3.1. Photomicrographs of human hippocampi depict the main neuropathologic endophenotypes analyzed in the current study. Hippocampal sclerosis (HS) is evaluated with H&E stain (panels A,C,E), whereas LATE-NC is operationalized with phospho-TDP-43 immunohistochemistry (IHC; panels B, D, and F). All photomicrographs depict mid-level hippocampal sections dissected in the coronal plane. Panels A and B show stained brain sections from a woman (APOE e3/e4) who died at age 83; autopsy revealed neither LATE-NC nor HS. Panels C and D are from a man (APOE e3/e4) who died at age 93 with LATE-NC Stage 2. Panels E and F are from a woman (APOE e3/e3) who died at age 95 with LATE-NC Stage 2 and comorbid HS. Note the relatively atrophic hippocampal profile in Panel E in comparison to A or C (same scale bar); the HS+ profile in panel E also demonstrates parenchymal rarefaction which can be appreciated even at low magnification. Phospho-TDP-43 immunoreactive intraneuronal inclusions are highlighted with arrows in panels D and F. The representative photomicrographs were from research participants of the University of Kentucky AD Research Center. Scale bar = 2mm in A, C, and E, 75 microns in B, D, and F.....52

Figure 3.2. Venn diagrams of the overlap between limbic-predominant age-related TDP-43 encephalopathy neuropathological change (LATE-NC) and hippocampal sclerosis (HS) cases in **A.** National Alzheimer's Coordinating Center (NACC) and **B.** Religious Orders Study and Rush Memory and Aging Project (ROSMAP). Only participants with non-missing case data for both LATE-NC and HS are included. LATE-NC = limbic-predominant age-related TDP-43 encephalopathy neuropathological change; HS = hippocampal sclerosis; NACC = National Alzheimer's Coordinating Center; ROSMAP = Religious Orders Study and Rush Memory and Aging Project.....53

Figure 3.3. Adjusted, meta-analytic, single nucleotide variant (SNV)-level p-values for hippocampal sclerosis (HS) and limbic-predominant age-related TDP-43 encephalopathy neuropathological change (LATE-NC) across **A.** *TMEM106B* ± 10kb and **B.** *GRN* ± 10kb. All analyses were adjusted for sex, age at death, cohort/study, and the first three genetic principal components. Horizontal dashed lines represent the Bonferroni-corrected thresholds for significance that account for the number of independent tests in each genomic region. A diamond represents the SNV with the smallest p-value. The previously identified *TMEM106B* SNV (Rutherford et al., 2012) is labeled and identified with an arrow. MOI = mode of inheritance; LATE-NC = limbic-predominant age-related TDP-43 encephalopathy neuropathological change; HS = hippocampal sclerosis.....54

Figure 3.4. Adjusted, meta-analytic, single nucleotide variant (SNV)-level p-values for hippocampal sclerosis (HS) and limbic-predominant age-related TDP-43 encephalopathy neuropathological change (LATE-NC) across *APOE* ± 10kb. All analyses were adjusted for sex, age at death, cohort/study, and the first three genetic principal components. The horizontal dashed line represents the Bonferroni-corrected threshold for significance that accounts for the number of independent tests in the *APOE* ± 10kb region. A diamond represents the SNV with the smallest p-value. MOI = mode of inheritance; LATE-NC = limbic-predominant age-related TDP-43 encephalopathy neuropathological change; HS = hippocampal sclerosis.....55

Figure 3.5. Adjusted, meta-analytic, single nucleotide variant (SNV)-level p-values for hippocampal sclerosis (HS) and limbic-predominant age-related TDP-43 encephalopathy neuropathological change (LATE-NC) across *ABCC9* ± 10kb assuming a recessive mode of inheritance (MOI). A recessive MOI was assumed for *ABCC9* since it has consistently been the MOI with the strongest HS association for *ABCC9* (Nelson et al., 2014; Nelson et al., 2015; Katsumata et al., 2017). All analyses were for sex, age at death, cohort/study, and the first three genetic principal components. The horizontal dashed line represents the Bonferroni-corrected threshold for significance that accounts for the number of independent tests in the *ABCC9* ± 10kb region. A diamond represents the SNV with the smallest p-value. The previously identified *ABCC9* SNV (Nelson et al., 2014) is labeled and identified with an arrow. MOI = mode of inheritance; LATE-NC = limbic-predominant age-related TDP-43 encephalopathy neuropathological change; HS = hippocampal sclerosis.....56

Figure 3.6. Expression quantitative trait loci (eQTL) analyses for rs1914361 and *ABCC9* gene expression across human tissues in the Genotype-Tissue Expression (GTEx) database. **A.** Multi-tissue eQTL plot of rs1914361 and *ABCC9* gene expression; **B.** *ABCC9* normalized gene expression stratified by rs1914361 minor alleles in the nucleus accumbens region of the brain; and **C.** *ABCC9* normalized gene expression stratified by rs1914361 minor alleles in the aorta region of the artery. GTEx = Genotype-Tissue Expression; NES = normalize effect size; eQTL = expression quantitative trait loci.....57

Figure 3.7. Adjusted odds ratio estimates and 95% confidence intervals for genetic single nucleotide variants (SNV) and *APOE* $\epsilon 4$ carrier status from separate regression models of hippocampal sclerosis (HS) fit using data from the National Alzheimer's Coordinating Center (NACC), the Religious Orders Study and Rush Memory and Aging Project (ROSMAP), and the meta-analysis of NACC and ROSMAP. All regression models were adjusted for sex, age at death, cohort/study, and the first three genetic principal components. Some regression models were also adjusted for limbic-predominant age-related TDP-43 encephalopathy neuropathological change (LATE-NC) case status. LATE-NC = limbic-predominant age-related TDP-43 encephalopathy neuropathological change; NACC = National Alzheimer's Coordinating Center; ROSMAP = Religious Orders Study and Rush Memory and Aging Project.....58

Figure 3.8. *ABCC8* and *ABCC9* gene expression in various human tissues in the **A.** Genotype-Tissue Expression (GTEx) and **B.** Functional Annotation of Human Long Noncoding RNAs via Molecular Mapping (FANTOM5) databases. In GTEx, central nervous system (CNS) tissues included Brodmann (1909) area 24, Brodmann (1909) area 9, C1 segment of cervical spinal cord, amygdala, caudate nucleus, cerebellar hemisphere, cerebellum, cerebral cortex, hippocampus proper, hypothalamus, nucleus accumbens, pituitary gland, and substantia nigra; vascular/smooth muscle tissues included aorta, atrium auricular region, coronary artery, tibial artery, endocervix, esophagus muscularis mucosa, urinary bladder, and uterus; and other tissues included all other tissue types. In FANTOM5, CNS tissues included amygdala, brain, caudate nucleus, cerebellum, diencephalon, dorsal thalamus, globus pallidus, hippocampal formation, locus ceruleus, medulla oblongata, middle frontal gyrus, middle temporal gyrus, occipital cortex, occipital lobe, olfactory apparatus, parietal lobe, pituitary gland, putamen, spinal cord, and substantia nigra; vascular/smooth muscle tissue included artery, heart, heart left ventricle, left cardiac atrium, mitral valve, smooth muscle, tricuspid valve, and uterus; and other tissues included all other tissue types. GTEx = Genotype-Tissue Expression; FANTOM5 = Functional Annotation of Human Long Noncoding RNAs via Molecular Mapping; TPM = transcripts per million; CNS = central nervous system.....59

Figure 3.9. Diagrams depicting potential causal relationships between the genes under study with positive findings (*TMEM106B*, *ABCC9*, *GRN*, and *APOE*) and TDP-43 proteinopathy/limbic-predominant age-related TDP-43 encephalopathy (LATE), hippocampal sclerosis (HS), and Alzheimer's disease (AD). **A.** The candidate genes and their corresponding colors in the diagrams, **B.** a diagram of the current study's prima facie results, and **C.** a diagram showing hypothetical mechanistic pathways that are compatible with the findings of the current study, including how AD neuropathologic

changes (often linked to the APOE risk allele) may fit in with the current study's results. LATE = limbic-predominant age-related TDP-43 encephalopathy; HS = hippocampal sclerosis; AD = Alzheimer's disease.....60

Figure 4.1. The effect of the minor allele flipping algorithm on the resulting smoothed,.....81

Figure 4.2. Statistical power for simulations assuming 5 continuous phenotypes, 5 causal variants, and causal effect sizes simulated from a normal distribution with $\mu = 0$ and $\sigma_{causal} = 0.15$82

Figure 4.3. Statistical power for simulations assuming 5 continuous phenotypes, 5 causal variants, and causal effect sizes simulated from a normal distribution with $\mu = 0$ and $\sigma_{causal} = 0.25$84

CHAPTER 1. INTRODUCTION

1.1 Alzheimer's Disease

Dementia is a clinical state characterized by a chronic, acquired loss of cognitive abilities which affects 47 million people worldwide [1]. While several neurodegenerative diseases can cause a dementia-like syndrome, Alzheimer's disease (AD) is the most common cause and accounts for up to 56% of dementia cases at autopsy [2]. While the hallmarks of the AD brain are pathologic build-ups of the amyloid- β ($A\beta$) and tau proteins, the exact cause of AD is still unknown [3].

The gold standard for diagnosing AD is via neuropathological examination where both $A\beta$ and tau must both be present, though this can only be conducted at autopsy [4]. AD can also be diagnosed while a person is alive through clinical examination with or without biomarker evidence [5]. Clinical diagnosis of AD has been shown to be inaccurate, especially in the absence of additional biomarker data [6, 7].

It has become increasingly clear that most dementia cases are afflicted by more than one neurodegenerative disease [8, 9]. A recent study recent community-based cohort study looking at the prevalence of multiple proteinopathies in older adults found that all individuals had the presence of at least one of tau, amyloid- β ($A\beta$), α -synuclein, or TAR-DNA binding protein 43 (TDP-43) pathologies at autopsy and only 6.4% of individuals presented with only a single proteinopathy [10].

1.2 Other Neurodegenerative Diseases

Several non-AD neuropathologies – including hippocampal sclerosis (HS), limbic predominant age-related TDP-43 encephalopathy (LATE), and brain arteriolosclerosis (B-

ASC) – are difficult to clinically distinguish from AD which has led some researchers to refer to them as “AD mimics” [9, 11, 12]. LATE is disease entity characterized by TDP-43 proteinopathy and cognitive impairment in aged populations that is separate from frontotemporal lobar degeneration with TDP-43 inclusions [9, 13-15]. HS is a pathologic finding that commonly co-occurs with LATE that is characterized by selective neuronal loss and gliosis of the hippocampal formation out of proportion of any AD pathology [16, 17]. B-ASC is a common subtype of cerebral small vessel disease that is associated with substantial cognitive impairment [12].

Unlike AD, which becomes less prevalent in the oldest old, all of the AD mimics become increasingly prevalent in individuals of advanced age. Among individuals beyond age 80, HS and LATE pathologies have been observed in ~30% of individuals [9] and B-ASC has been identified in over 80% of individuals [12]. Since these AD mimics are highly prevalent and difficult to detect clinically, it increases the likelihood of misclassification when relying solely on clinical diagnoses of AD especially when the goal is to study the etiology of AD.

1.3 Genetics

Genome-wide association studies (GWAS) with large sample sizes have successfully identified dozens of genes and single nucleotide polymorphisms (SNP) associated with clinical AD [18-21]. Additionally, through the use of proxy AD cases which are defined via self-reports of familial AD, AD GWAS sample sizes have increased even more and are able to detect SNPs with much smaller effect sizes [22-24]. While many of the earlier

identified AD genes and SNPs have been corroborated in follow-up studies and external cohorts, several AD-associated genetic loci remain largely uncharacterized.

Risk genes and SNPs have also been identified for HS [9, 25-38], LATE [9, 26, 29, 32, 39-41], and B-ASC [12, 42-44]. Notably, there are several risk genes that are shared either among the mimics or with AD including *APOE*, *ABCC9*, *GRN*, and *TMEM106B*. While genetic pleiotropy has been observed in neurodegeneration [39, 45-49], it is still unclear if these genes actually cause more than one neuropathology or if their apparent associations are due misclassification of clinical phenotypes.

1.4 Dissertation Outline

This dissertation presents studies on the associations of several neurodegenerative pathologies and the *WWOX* gene, the gene-based associations of previously identified HS risk genes and LATE-NC, and the development of a new statistical method for testing gene-based, multi-phenotype associations. In Chapter 2, variants in a recently identified clinical AD risk locus encompassing the *WWOX* and *MAF* genes are tested for associations with AD-specific pathologies along with HS, LATE-NC, and B-ASC. The major findings from this study included significant associations with HS, LATE-NC, and B-ASC and, notably, no significant associations with AD-related pathologies. This suggests that the original clinical AD findings for the *WWOX/MAF* locus were driven by non-AD pathologies. In Chapter 3, never-before-analyzed HS and LATE-NC data are used to reevaluate previously identified HS risk genes and risk SNPs. The major findings from this study include both HS and LATE-NC being associated with the *APOE* and *TMEM106B* genes and only HS being associated with the *ABCC9* and *GRN* genes. Additionally, several novel SNPs in the *TMEM106B*, *ABCC9*, and *APOE* genes were identified and showed biological plausibility.

In Chapter 4, we use the framework of function-on-scalar regression models from the branch of statistics known as functional data analysis to development a new statistical test for gene-based, multi-phenotype associations. Our newly derived test performs similarly to another test based on function-on-scalar regression models and outperforms competing methods, especially when the sample sizes are smaller and the genetic associations weaker. The conclusion of the dissertation is discussed in Chapter 5.

CHAPTER 2. ASSOCIATION BETWEEN *WWOX/MAF* VARIANTS AND DEMENTIA-RELATED NEUROPATHOLOGIC ENDOPHENOTYPES

2.1 Abstract

The genetic locus containing the *WWOX* and *MAF* genes was implicated as a clinical Alzheimer's disease (AD) risk locus in two recent large meta-analytic genome wide association studies (GWAS). In a prior GWAS, we identified a variant in *WWOX* as a suggestive risk allele for hippocampal sclerosis (HS). We hypothesized that the *WWOX/MAF* locus may be preferentially associated with non-plaque- and non-tau-related neuropathological changes (NC). Data from research participants with GWAS and autopsy measures from the National Alzheimer's Coordinating Center (NACC) and the Religious Orders Study and Memory and the Rush Aging Project (ROSMAP) were meta-analyzed. Notably, no variants in the locus were significantly associated with AD-related NC. However, several *WWOX/MAF* variants had significant adjusted associations with limbic-predominant age-related TDP-43 encephalopathy NC (LATE-NC) and HS. The LATE-NC and HS associations remained unchanged after adjustment for AD-associated neurofibrillary pathology and, separately, for neuritic amyloid plaques, suggesting that these associations are independent of the presence or severity of AD pathology.

2.2 Introduction

The human WW domain-containing oxidoreductase (*WWOX*) and MAF bZIP transcription factor (*MAF*) genes are situated close to each other on chromosome 16q23. The normal functions of these genes are incompletely characterized. *WWOX* protein plays roles in transcription regulation, glucose metabolism, and central nervous system

development [50], while the protein encoded by the *MAF* gene is a transcription factor that regulates cellular processes including T-cell susceptibility to apoptosis.

This locus has been implicated in human disease. *WWOX* has been hypothesized to play a role in neurodegenerative disease, particularly AD [51-53]. A large genome-wide association study (GWAS) of clinical AD suggested that *WWOX* confers AD risk in non-Hispanic White individuals [18] and a follow-up GWAS in African American individuals nominally replicated this association [19]. More recently, the largest AD GWAS to date also found an association between AD and *MAF*, the gene just downstream of *WWOX* [54]. In addition to AD, *WWOX* has also shown suggestive linkage with autism and schizophrenia [55, 56] and *MAF* has been associated with thyroid-related diseases, such as Graves' disease and Hashimoto's disease [57]. Nonetheless, the neurochemistry of *WWOX* and *MAF* in the human brain, and in human disease, is still poorly understood.

It has become increasingly clear that AD and related dementias (ADRD) are highly complex at both the individual level (multiple pathologies per person) and in a population (many different combinations of mixed pathologies). Thus, multiple neuropathological changes are associated with the AD clinical syndrome and these neurodegenerative diseases often co-occur, especially in older age [8, 9]. A recent community-based cohort study looking at the prevalence of multiple proteinopathies in older adults found that all individuals had the presence of at least one of tau, amyloid- β ($A\beta$), α -synuclein, or TAR-DNA binding protein 43 (TDP-43) pathologies at autopsy and only 6.4% of individuals presented with only a single proteinopathy [10].

A previous GWAS found *WWOX* to be a gene suggestive for association with hippocampal sclerosis (HS) pathology [28]. The brain conditions which were previously

referred to as “HS-Aging” and “HS dementia” are now subsumed under a broader disease category, and are characterized by the presence of comorbid TDP-43 proteinopathy, which is a more sensitive and specific feature. The condition was recently classified with the term limbic-predominant age-related TDP-43 encephalopathy (LATE) [9]. The presence of the neuropathological changes underlying LATE (LATE-NC) is associated with a dementia syndrome similar to AD [9].

Given that we had found a suggestive link between *WWOX* variants and HS, and others found an association between the *WWOX/MAF* locus and clinical AD, we hypothesized that a more definitive conclusion could be reached via a pathology-based study of separate cohorts with both genetics and pathologic information (including TDP-43 proteinopathy) available. We investigated whether the *WWOX/MAF* AD association could be due to neuropathological changes other than AD-type pathological hallmarks, amyloid plaques and neurofibrillary tangles. GWAS data and autopsy-confirmed neuropathological endophenotypes were gathered from the National Alzheimer’s Coordinating Center (NACC) and from the Religious Orders Study and the Rush Memory and Aging Project (ROSMAP), to resolve novel associations between pathological findings and *WWOX/MAF* genetic variation.

2.3 Materials and Methods

2.3.1 Study Participants

Phenotypic data from NACC were linked with genotype data from the Alzheimer’s Disease Genetics Consortium (ADGC). Individuals who died at age 65 years or older were included in this study. Similar to other studies using NACC data [58], individuals were

excluded from the NACC cohort if at least one of 19 rare brain diseases were diagnosed (**Supplemental Table 2.1**) or if they were missing any adjustment variables or all of the endophenotypes under study.

The ROSMAP study has been described in detail elsewhere [59]. Briefly, data were acquired from two well-characterized cohort studies of aging and dementia. The Religious Orders Study (ROS), begun in 1994, and the Rush Memory and Aging Project (MAP), begun in 1997, involve older adults who enrolled without dementia, agreed to annual clinical evaluations and organ donation at death, and signed an Anatomical Gift Act for brain donation. Written informed consent was obtained from participants, and research was carried out in accordance with Institutional Review Board (IRB)-approved protocols. ROSMAP data are available online at the Rush Alzheimer's Disease Center Resource Sharing Hub (<https://www.radc.rush.edu/>), as well as on the Accelerating Medicines Partnership-Alzheimer's Disease (AMP-AD) Knowledge Portal (syn3219045).

2.3.2 Neuropathological Endophenotype Definitions

In the NACC Neuropathology (NP) dataset, LATE-NC was defined as either present or absent using the “distribution of TDP-43 immunoreactive inclusions” variables indicating if TDP-43 proteinopathy was observed in either the hippocampus (NPTDPC NACC field), entorhinal/inferior temporal cortex (NPTDPD), or neocortex (NPTDPE) in a case lacking overall diagnosis of frontotemporal lobar degeneration (FTLD)-TDP. HS was defined as either present or absent based on the “hippocampal sclerosis of CA1 and/or subiculum” (NPHIPSCL) variable using the “unilateral,” “bilateral,” and “present but laterality not assessed” response categories. Arteriolosclerosis was defined similarly using

the “arteriolosclerosis” (NACCARTE) variable and collapsing the “moderate” and “severe” response categories. Presence of neurofibrillary tangles was also defined dichotomously using the “Braak stage for neurofibrillary degeneration (B score)” (NACCBRAA) variable and collapsing the “stage V (B3)” and “stage VI (B3)” response categories. Presence of neuritic plaques was defined dichotomously using the “frequent neuritic plaques (C3)” response category of the “density of neocortical neuritic plaques (CERAD score) (C score)” (NACCNEUR) variable.

In ROSMAP, LATE-NC was defined dichotomously using the “TDP-43 stage” (tdp_st4) variable and collapsing the 2nd and 3rd stages in cases lacking FTLD-TDP. HS was defined dichotomously by the “hippocampal sclerosis was rated as definitely present with CA1 region affected” response category of the “definite presence of typical hippocampal sclerosis” (hspath_typ) variable. Arteriolosclerosis was defined dichotomously using the “arteriolosclerosis” (arteriol_scler) variable and collapsing the “moderate” and “severe” response categories. Presence of neurofibrillary tangles was defined dichotomously using the “semiquantitative measure of neurofibrillary tangles” (braaksc) variable and collapsing the “V” and “VI” response categories. Presence of neuritic plaques was defined dichotomously by the “definite” response category of the “semiquantitative measure of neuritic plaques” (ceradsc) variable.

2.3.3 Quality Control of Genotype Data

For NACC participants, genomic data from the ADGC imputed using the Haplotype Reference Consortium (ADGC-HRC) were used [60]. The genetic data for ROSMAP was also imputed using the HRC and the methods have been described in detail

elsewhere [61]. Standard GWAS quality control (QC) procedures were performed separately on the ADGC and ROSMAP genotype data using PLINK1.9 [62, 63]. Variants were excluded if they were missing in more than 5% of samples, if they had a minor allele frequency less than 1%, or if they had Hardy-Weinberg Equilibrium (HWE) p-values $< 1 \times 10^{-6}$. Individuals were excluded if they were missing more than 5% of genotypes. Two individuals were considered related if they had an identity by descent measure of at least 0.25, which indicates that they are second-degree relatives. For related pairs, the individual with the lowest call rate was excluded.

NACC and ROSMAP genotype data were separately merged with 1000 Genomes data Phase 3. Principal components (PCs) were calculated for the merged data sets using the “pca” procedure in PLINK1.9, and the first two PCs were plotted. Using the known ancestries from the 1000 Genomes data, individuals of European ancestry in the ADGC-HRC and ROSMAP data sets were identified and all other individuals were excluded from the analysis.

2.3.4 Variant-Level Associations

The gene boundaries of *WWOX* and *MAF* were defined based on the canonical transcripts (*WWOX*: 78,133,309 – 79,246,564; *MAF*: 79,627,744 – 79,634,622) using the GRCh37/hg19 gene range list from PLINK (<https://www.cog-genomics.org/plink/1.9/resources>). Since both of the genome-wide significant *WWOX/MAF* AD variants (rs62039712 from Kunkle et al. and rs450674 from Bellenguez et al.) were in the intergenic region between the *WWOX* and *MAF* genes, we defined the *WWOX/MAF* locus to be from 78,133,309 to 79,634,622 +/- 250kb of flanking.

Associations between each endophenotype and each variant were tested separately in the NACC and ROSMAP datasets using binary logistic regression models assuming each of the three most commonly used modes of inheritance (MOI): additive, dominant, and recessive. Variants were excluded from the analyses if they were multiallelic or if there were fewer than 15 minor alleles present across all participants. All regression models were fit in R programming language [64], version 4.0.4, using the glm function and adjusted for age at death, sex, ADGC data selection round (for NACC data) or ROS/MAP study (for ROSMAP data), and the first three PCs. Odds ratios (OR) were calculated for each variant by exponentiating the variant's beta estimate. Since some endophenotypes were only available in a subset of participants, PCs were calculated separately for each endophenotype using the "pca" procedure in PLINK1.9. NACC and ROSMAP variant-level results were meta-analyzed using a fixed-effect, inverse-variance meta-analysis via the metagen function from the meta R package, version 4.18-0 [65]. Plots of cohort-specific and meta-analyzed variant-level p-values were created using LocusZoom Standalone, version 1.4 (https://genome.sph.umich.edu/wiki/LocusZoom_Standalone) [66]. Linkage disequilibrium estimates were computed using LDlink assuming a CEU population (<https://ldlink.nci.nih.gov/>) [67].

2.3.5 Variant Prioritization and Downstream Analyses

Statistically significant variants were identified using a Bonferroni-corrected threshold for significance that accounts for the effective number of independent tests in the region. The effective number of independent tests in the region was calculated for each endophenotype using the method of Gao et al. [68]. Briefly, Pearson's correlation

coefficient was calculated for all pairs of variants and these coefficients were placed in a square matrix. The eigenvalues of the matrix were then computed and ordered from largest to smallest and the effective number of independent tests was defined to be the smallest number of ordered eigenvalues that account for 99.5% of the sum of all eigenvalues. The Bonferroni-corrected threshold for identifying prioritized variants for an endophenotype was defined as a variant-level p-value less than 0.05 divided by the effective number of independent tests in the region for the endophenotype.

Prioritized variants were investigated for expression quantitative trait loci (eQTL) and splicing quantitative trait loci (sQTL) associations using the Genotype-Tissue Expression (GTEx) Project's V8 public data [69] and the BRAINEAC Brain eQTL Almanac (<http://braineac.org/>) [70]. Prioritized variants were also investigated for associations with other molecular mechanisms using INFERNO software [71].

2.3.6 Sensitivity Analyses

The dependency of the study's results on several analytic choices were investigated. In addition to including 250kb of flanking on both sides of the *WWOX/MAF* locus, all analyses were conducted assuming no flanking and 25kb of flanking. Since some neurodegenerative diseases are more pronounced at later ages and some variant effects may be age-dependent or only affect the age of onset, all analyses were also conducted on the subset of individuals with ages of death of 75 years or older. Finally, to determine if significant variant-level results were independent of ADNC, all analyses were also conducted while adjusting for the presence of neurofibrillary tangles and, separately, for the presence of neuritic plaques.

2.4 Results

In the NACC data set, $n=3,749$ individuals had available data for at least one of the endophenotypes along with GWAS data. In ROSMAP, a total of $n=1,390$ individuals had available data for at least one of the endophenotypes along with GWAS data. **Table 2.1** shows a summary of individual characteristics and endophenotypes for both NACC and ROSMAP participants. NACC participants with neurofibrillary tangles ($p<0.001$), neuritic plaques ($p<0.001$), and arteriolosclerosis ($p<0.001$) tended to be younger at death. Conversely, ROSMAP participants with an endophenotype present tended to be older at death and were less likely to be male (all $p<0.05$).

2.4.1 Variant-Level Associations

A total of 9,492 genetic variants in the *WWOX/MAF* locus passed QC in NACC and 8,953 variants passed QC in ROSMAP. A total of 8,256 variants were shared between NACC and ROSMAP and were included in the meta-analysis. Notably, rs62039712, the top *WWOX/MAF* variant from the Kunkle et al. clinical AD GWAS [5], did not pass QC in either data set because it was missing in greater than 5% of individuals. No variants were in high enough linkage disequilibrium with rs62039712 in a CEU populations in LDlink to serve as proxies (no variants with $R^2 > 0.4$ within 500kb of rs62039712).

The *WWOX* variant previously found to be genome-wide suggestive for HS with a recessive MOI, rs55751884 [14], had nominally significant adjusted associations with HS ($p=0.01133$) in NACC and neurofibrillary tangles in both NACC and ROSMAP (NACC: $p=0.00330$; ROSMAP: $p=0.03416$) assuming a recessive MOI (**Table 2.2**). However,

while the NACC and ROSMAP odds ratios for rs55751884 on neurofibrillary tangles were of similar magnitude, they did not point in the same direction despite having the same minor allele (NACC: OR=2.38; ROSMAP: OR=0.32) and the adjusted meta-analytic p-value did not reach nominal significance ($p=0.10566$). Notably, the adjusted association between rs55751884 and HS remained nominally significant when restricted to participants not included in the 2014 HS GWAS (meta-analysis $p=0.02568$) representing a nominal replication of that HS association (**Supplemental Table 2.2**). When additive and dominant MOIs were assumed, additional adjusted associations reached nominal statistical significance including brain arteriolosclerosis in NACC (dominant MOI $p=0.03008$) (**Supplemental Table 2.2**).

The recently identified genome-wide significant clinical AD risk variant near the *MAF* gene, rs450674 which is between *WWOX* and *MAF* approximately 53kb away from *MAF*'s 3' end [7], had a nominally significant adjusted association with neurofibrillary tangles in NACC ($p=0.00637$) and in the meta-analysis of NACC and ROSMAP ($p=0.03227$) assuming an additive MOI (**Table 2.2**). No additional adjusted associations were found to be nominally significant when assuming a recessive or dominant MOI (**Supplemental Table 2.2**).

2.4.2 Variant Prioritization and Downstream Analyses

The largest estimate of the effective number of independent tests for the *WWOX/MAF* locus ± 250 kb was 1,364 in NACC and 804 in ROSMAP. The larger of these two estimates was used to compute the Bonferroni-corrected threshold for the

WVOX/MAF locus +/- 250kb of 3.67×10^{-5} (0.05/1,364). Variants with p-values less than this threshold were prioritized for further investigation.

Associations with ADNC endophenotypes were notably absent for the recently identified clinical AD loci of rs62039712 and rs450674 (**Figure 2.1**). However, a locus centered around the rs11640136 variant near the *CLEC3A* gene contained several variants that surpassed the Bonferroni-corrected threshold for significance for the *WVOX/MAF* locus \pm 250kb in NACC for neuritic plaques (Supplemental Figure 1). This trend is absent in the ROSMAP cohort. There is a similar trend in the same locus for neurofibrillary tangles in NACC, but no variants surpass the Bonferroni-corrected threshold for significance for the *WVOX/MAF* locus \pm 250kb. The two top variants for the neuritic plaques (rs11640136) and neurofibrillary tangles (rs12922846) loci in NACC are not in linkage disequilibrium with one another ($R^2=0.015$ in a CEU population) nor are they in linkage disequilibrium with the previously identified HS locus nearby (both $R^2 < 0.015$ in a CEU population). The association of the variants near the *CLEC3A* gene and neuritic plaques are not significant in a meta-analysis of NACC and ROSMAP (**Figure 2.1**).

In NACC, 32 unique variants met the Bonferroni-corrected significance threshold for the *WVOX/MAF* region \pm 250kb for at least one endophenotype and one MOI. Ten of those variants were associated with LATE-NC, 10 were associated with HS, and 12 were associated with neuritic plaques. None of the variants were associated with neurofibrillary tangles or brain arteriolosclerosis. LATE-NC was most strongly associated with the rs8052915 locus (recessive MOI $p=4.02 \times 10^{-6}$), HS was most strongly associated with the rs10438625 locus (additive MOI $p=7.56 \times 10^{-6}$), and neuritic plaques were most strongly associated with the rs112959604 locus (dominant MOI $p=8.63 \times 10^{-6}$). The HS locus,

centered around rs10438625, is located approximately 85kb downstream of an earlier identified *WWOX* HS risk variant, rs55751884 [14], though the two variants are not in linkage disequilibrium ($R^2 < 0.001$ in a CEU population). Interestingly, the neuritic plaques locus, rs112959604, is located in the intergenic region between *CLEC3A* and the 5' end of *WWOX* and is in linkage disequilibrium with variants within the *CLEC3A* gene (**Figure 2.1**). See **Supplemental Table 2.3** for additional details on the prioritized variants.

In ROSMAP, 20 unique variants met the Bonferroni-corrected significance threshold for the *WWOX/MAF* region ± 250 kb for at least one endophenotype and one MOI. Unlike NACC, all of these variants were associated with brain arteriolosclerosis and were centered around the rs79815901 locus ($p=3.00 \times 10^{-8}$). The rs79815901 variant is located approximately 90kb upstream of the earlier identified *WWOX* HS variant, rs55751884 [14], though these two variants are not in linkage disequilibrium ($R^2 < 0.001$ in a CEU population). Additionally, rs79815901 is approximately 5kb downstream of the HS locus identified in NACC centered around rs11150053, but these two variants are also not in linkage disequilibrium ($R^2 = 0.055$ in a CEU population). See **Supplemental Table 2.3** for additional details on the prioritized variants.

Eight thousand two hundred and fifty-six variants were shared between NACC and ROSMAP in the *WWOX/MAF* locus ± 250 kb and were meta-analyzed across NACC and ROSMAP. Five variants had meta-analytic p-values that met the Bonferroni-corrected significance threshold for the *WWOX/MAF* region ± 250 kb for at least one endophenotype and one MOI. Two of these variants, rs6564590 and rs7404901, were associated with LATE-NC assuming an additive MOI ($p=1.07 \times 10^{-5}$ and $p=1.56 \times 10^{-5}$, respectively). Both variants are located in the same region of *WWOX* as the LATE-NC locus identified in the

NACC-only analysis centered around rs8052915 and are in relatively high linkage disequilibrium with rs8052915 (rs6564590: $R^2 = 0.765$; rs7404901: $R^2 = 0.551$). Two additional variants, rs9925100 and rs9930659, were associated with HS while assuming a recessive MOI ($p=1.34 \times 10^{-5}$ and $p=1.82 \times 10^{-5}$, respectively). These two variants are in linkage disequilibrium with one another ($R^2 = 0.726$ in a CEU population) and are located approximately 856kb downstream from the HS locus identified in the NACC-only analysis centered around rs11150053. The remaining variant, rs4435266, was associated with brain arteriolosclerosis while assuming a dominant MOI ($p=2.02 \times 10^{-5}$) (**Table 2.3** and **Figure 2.2**). While the other endophenotypes were associated with variants in either the NACC-only or the ROSMAP-only analyses, they did not have any associations with meta-analytic p-values that met the Bonferroni-corrected significance threshold (Supplemental Figure 2).

None of the prioritized meta-analytic variants were found to be associated with eQTLs or sQTLs for *WWOX*, *MAF*, or any other proximal genes in GTEx. However, all three variants were found to have notable associations in BRAINEAC. The two HS variants, rs9925100 and rs9930659, had nominally significant eQTL associations for *WWOX* (both brain tissue-wide p-values $< 3.9 \times 10^{-3}$) with the hippocampus and putamen regions having the strongest single-tissue associations. Both of the LATE-NC variants, rs6564590 and rs7404901, had nominally significant eQTL associations with *MAF* (brain tissue-wide $p=0.040$ and $p=0.012$, respectively), with the thalamus region having the strongest single-tissue association for rs6564590 ($p=0.0096$) and the frontal cortex region having the strongest single-tissue association for rs7404901 ($p=0.0036$). The brain arteriolosclerosis variant, rs4435266, also had a nominally significant eQTL association with *MAF* (brain tissue-wide $p=0.019$) with the cerebellum region having the strongest

single-tissue association ($p=0.0088$). Additionally, in INFERNO the LATE-NC variants were found to be eQTLs for Roadmap enhancers in the blood and immune organ tissues, the brain arteriolosclerosis variant was found to be an eQTL for Roadmap enhancers in blood and skeletal muscle tissues, and the HS variants were found to be eQTLs for both Roadmap and FANTOM5 enhancers in the blood and Roadmap enhancers in immune organ and skeletal muscle tissues.

2.4.3 Sensitivity Analyses

2.4.3.1 Varying *WWOX/MAF* Flanking

The Bonferroni-corrected threshold for the *WWOX/MAF* locus was estimated to be 3.12×10^{-5} (0.05/1,214) with 25kb of flanking and 4.19×10^{-5} (0.05/1,194) with 0kb of flanking. Not surprisingly, all five of the prioritized meta-analytic variants along with the majority of the NACC-only and ROSMAP-only variants identified in the primary analysis with 250kb of flanking were also identified when the flanking was reduced. The exception to this were the variants associated with neuritic plaques in the NACC-only analysis which were located just upstream of *WWOX* near the *CLEC3A* gene. Of the 12 variants found to be associated with neuritic plaques when the *WWOX/MAF* locus was flanked by 250kb, only one variant (rs79416778) remained with no flanking. This result highlights the specificity of the neuritic plaques association with the *CLEC3A* gene over the *WWOX* gene. See **Supplemental Table 2.4** for the complete results.

2.4.3.2 Age of Death 75+

The odds ratio estimates for all three of the prioritized meta-analytic variants remained largely unchanged when the analyses were restricted to only those individuals with an age of death of 75 years of age or older. Additionally, the meta-analytic p-values for the associations between rs6564590 and LATE-NC ($p=2.85 \times 10^{-5}$) and rs4435266 and brain arteriolosclerosis ($p=3.07 \times 10^{-5}$) and remained below the Bonferroni-corrected threshold. The remaining associations were all nominally significant, but did not quite meet the Bonferroni-corrected threshold (all remaining p-values $\leq 6.12 \times 10^{-5}$). These findings suggest that age of death does not meaningfully impact the associations between the prioritized variants and the endophenotypes. See **Supplemental Table 2.5** for the complete results.

2.4.3.3 Adjusting for ADNC

The odds ratio estimates for all five of the prioritized meta-analytic variants remained largely unchanged when the analyses were adjusted for neurofibrillary tangles and, separately, neuritic plaques. Additionally, all of the meta-analytic p-values remained below the Bonferroni-corrected threshold when adjusted for neuritic plaques and all but one remained below the threshold when adjusted neurofibrillary tangles (the exception was rs7404901 where $p=5.46 \times 10^{-5}$). These findings suggest that the associations of the prioritized variants are independent of ADNC. See **Supplemental Table 2.6** for the complete results.

2.5 Discussion

Using autopsy-confirmed neuropathologic endophenotypes, we evaluated the genetic associations between the *WWOX/MAF* locus and several neurodegenerative diseases using neuropathological changes to operationalize the presence and severity of the diseases. We found significant adjusted meta-analytic associations between *WWOX* variants and LATE-NC, HS, and brain arteriolosclerosis. While previous GWASs linked variants in the *WWOX/MAF* locus with HS and clinical AD, the associations with LATE-NC and brain arteriolosclerosis have never been reported. Furthermore, since these associations remained significant after adjustment for AD-related neuropathological changes, it suggests that the LATE, HS, and brain arteriolosclerosis neuropathological changes associated with *WWOX/MAF* are independent of ADNC.

The novel neuritic plaque signal found in NACC near the *CLEC3A* gene is intriguing since other *CLEC* family genes have been linked to AD and inflammation [72-74]. Additionally, a recent genome-wide interaction analysis found evidence of variant-by-variant interactions for neurofibrillary tangles involving variants near *CLEC3A* and *WWOX* [75]. Further investigations into the influence of *CLEC3A* on neurodegenerative disease are warranted.

The previously identified AD-associated *WWOX/MAF* variant, rs62039712, did not pass QC in either dataset and did not have any proxy variants, and was only available in two of the 12 Stage 1 GWAS cohorts from Kunkle et al. These factors indicate that rs62039712 was difficult to impute. The recently identified AD-associated *MAF* variant, rs450674 [54], which is located approximately 362kb downstream of *WWOX* and is not in linkage disequilibrium with rs62039712, was found to have a nominally significant

adjusted association with neurofibrillary tangles in NACC ($p=0.00637$) and in the meta-analysis of NACC and ROSMAP ($p=0.03620$). However, since Kunkle et al., Bellenguez et al., and the present study have all utilized data from the ADGC – albeit using different, more specific phenotypes in the case of the current study – this region of the genome should be investigated further in other datasets to better understand its influence on AD risk.

We examined the associations of the genome-wide suggestive HS *WWOX* variant, rs55751884, with neuropathological endophenotypes. The rs55751884 variant was nominally significant in adjusted association tests for neurofibrillary tangles in both NACC and ROSMAP (NACC: $p=0.00330$; ROSMAP: $p=0.03416$), HS in NACC ($p=0.01133$), and borderline significant for neuritic plaques in NACC ($p=0.08202$). Even though the meta-analytic association between HS and rs55751884 did not reach the Bonferroni-corrected threshold for significance in our current study, that same region of *WWOX* had the strongest association with arteriolosclerosis in ROSMAP, which also merits additional investigation.

Given the abundant evidence that mixed pathologies are highly prevalent in elderly populations, the hypothesis that *WWOX* is associated with several neuropathological endophenotypes fits in with recent studies looking at genetic pleiotropy in neurological conditions [39]. Pleiotropic effects have been found between AD and Parkinson’s disease [45], AD and amyotrophic lateral sclerosis [46], early-onset AD and frontotemporal dementia [47], AD-related psychosis and schizophrenia [48], and LATE-NC and FTLDD-TDP [9]. A specific example is the *MAPT* gene which is a risk allele for many tauopathies, and also for Parkinson’s disease (not a condition linked to tau pathology) [49]. Pleiotropic effects have also been found between AD-related neuropathological changes like neuritic

plaques, neurofibrillary tangles, and cerebral amyloid angiopathy [76]. Further, it has been shown that brain arteriolosclerosis is linked to HS and LATE-NC [12, 77]. Our data indicate that the associations between the *WWOX/MAF* locus and LATE-NC, HS, and brain arteriolosclerosis were independent of ADNC. Thus, *WWOX* is apparently associated with more than one clinico-pathologic entity. Since *WWOX* is also known to play a role in molecular functions [78], autism spectrum disorder [55], multiple sclerosis [79], schizophrenia [56], and brain volume [80], it is a good target for additional follow-up studies.

There are limitations to our study. Because data come from studies employing variable study designs and are highly homogeneous, the degree to which findings are generalizable is unknown, especially with respect to individuals of non-Caucasian ancestries. These suggestive findings extend prior research in the field that linked the *WWOX/MAF* locus with neurodegenerative phenotypes. Yet these findings need corroborative evaluations in additional data sets to evaluate the relationships between genetics and neuropathologic data.

2.6 Conclusion

In conclusion, we showed using large genetic datasets and autopsy-derived endophenotypes that neuropathological endophenotypes related to LATE, HS, and brain arteriolosclerosis were associated with *WWOX/MAF* gene variants. While clinical diagnoses of AD may be helpful for discovering dementia-related genetic variation, our study adds to the growing body of literature highlighting the complexity of dementia phenotypes, and the benefit of leveraging autopsy-derived data for studies of aging-related brain disease.

Table 2.1. Individual characteristics stratified by endophenotype status for National Alzheimer's Coordinating Center (NACC) and Religious Orders Study and Rush Memory and Aging Project (ROSMAP) participants.

Endophenotype Status	NACC			ROSMAP		
	Number of Participants (%)	Age at Death, Mean (SD)	Female, N (%)	Number of Participants (%)	Age at Death, Mean (SD)	Female, N (%)
Hippocampal Sclerosis	N=631	85.90 (8.30)	319 (50.6)	N=1200	89.55 (6.47)	812 (67.7)
Absent	542 (85.9)	85.89 (8.43)	270 (49.8)	1091 (90.9)	89.27 (6.45)	729 (66.8)
Present	89 (14.1)	85.99 (7.50)	49 (55.1)	109 (9.1)	92.38 (6.03)	83 (76.1)
LATE-NC	N=412	85.08 (7.86)	207 (50.2)	N=1130	89.83 (6.43)	775 (68.6)
Absent	291 (70.6)	84.93 (8.08)	138 (47.4)	733 (64.9)	88.75 (6.57)	471 (64.3)
Present	121 (29.4)	85.44 (7.33)	69 (57.0)	397 (35.1)	91.81 (5.64)	304 (76.6)
Neurofibrillary Tangles	N=3760	82.46 (8.23)	1939 (51.6)	N=1390	89.43 (6.54)	944 (67.9)
Braak Stage 0 to IV	1236 (32.9)	84.97 (8.39)	639 (51.7)	1046 (75.3)	88.91 (6.73)	679 (64.9)
Braak Stage V or VI	2524 (67.1)	81.24 (7.87)	1300 (51.5)	344 (24.7)	91.00 (5.63)	265 (77.0)
Neuritic Plaques	N=3764	82.48 (8.23)	1940 (51.5)	N=1222	89.54 (6.51)	825 (67.5)
None/Sparse/Moderate	1269 (33.7)	85.91 (8.28)	612 (48.2)	810 (66.3)	89.23 (6.80)	510 (63.0)
Frequent	2495 (66.3)	80.73 (7.64)	1328 (53.2)	412 (33.7)	90.15 (5.84)	315 (76.5)
Brain Arteriolosclerosis	N=2999	82.94 (8.30)	1514 (50.5)	N=1390	89.43 (6.54)	944 (67.9)
None/Mild	1720 (57.4)	81.93 (8.38)	832 (48.4)	1013 (72.9)	88.99 (6.43)	666 (65.7)
Moderate/Severe	1279 (42.6)	84.30 (8.00)	682 (53.3)	377 (27.1)	90.62 (6.67)	278 (73.7)

NACC = National Alzheimer's Coordinating Center; ROSMAP = Religious Orders Study and Rush Memory and Aging Project; SD = standard deviation; LATE-NC = limbic-predominant age-related TDP-43 encephalopathy neuropathological changes.

Table 2.2. Adjusted results for previously published HS and clinical AD variants in the WWOX/MAF locus. All analyses adjusted for age at death, sex, Alzheimer's Disease Genetics Consortium (ADGC) cohort or Religious Orders Study and Memory and Aging Project (ROSMAP) study, and the first three genetic principal components. The rs55751884 results are reported assuming a recessive mode of inheritance (MOI) as that was the MOI with the strongest association in Nelson et al., 2014. The rs450674 results are reported assuming an additive MOI as that was the MOI reported in Bellenguez et al., 2020.

Variant	MOI	Endophenotype	NACC			ROSMAP			Meta-Analysis		
			OR	95% CI	P-value	OR	95% CI	P-value	OR	95% CI	P-value
rs55751884	Rec.	Hippocampal Sclerosis	5.53	1.51-19.51	0.01133	1.88	0.53-5.21	0.29671	3.04	1.32-6.97	0.00878
		LATE-NC	3.69	0.75-20.26	0.10571	1.21	0.48-2.92	0.67826	1.58	0.73-3.44	0.24823
		Neurofibrillary Tangles	1.66	0.94-3.07	0.08202	0.57	0.21-1.38	0.22405	1.23	0.75-2.02	0.41942
		Neuritic Plaques	2.38	1.32-4.57	0.00330	0.32	0.07-0.93	0.03506	1.58	0.91-2.74	0.10566
		Brain Arteriolosclerosis	1.24	0.73-2.08	0.42069	1.48	0.63-3.32	0.35418	1.30	0.84-2.02	0.23621
rs450674	Add.	Hippocampal Sclerosis	0.77	0.48-1.23	0.26429	1.29	0.85-2.00	0.23481	1.02	0.74-1.39	0.91268
		LATE-NC	1.05	0.76-1.46	0.75143	0.97	0.81-1.17	0.75133	0.99	0.84-1.16	0.90518
		Neurofibrillary Tangles	0.93	0.83-1.04	0.20840	0.95	0.79-1.13	0.56609	0.94	0.85-1.03	0.16999
		Neuritic Plaques	0.86	0.78-0.96	0.00637	1.05	0.87-1.27	0.62292	0.90	0.82-0.99	0.03227
		Brain Arteriolosclerosis	0.93	0.84-1.04	0.22636	1.04	0.87-1.25	0.66185	0.96	0.87-1.06	0.41977

NACC = National Alzheimer's Coordinating Center; ROSMAP = Religious Orders Study and Rush Memory and Aging Project; LATE-NC = limbic-predominant age-related TDP-43 encephalopathy neuropathological changes; MOI = mode of inheritance; Rec. = recessive MOI; Add. = additive MOI; OR = odds ratio; and CI = confidence interval.

Table 2.3. Variant-level results for variants with uncorrected meta-analytic p-values that met the Bonferroni-corrected threshold for significance for the WWOX/MAF locus \pm 250kb. All analyses adjusted for age at death, sex, Alzheimer's Disease Genetics Consortium (ADGC) cohort or Religious Orders Study and Memory and Aging Project (ROSMAP) study, and the first three genetic principal components.

Endo-phenotype	Variant	MOI	NACC			ROSMAP			Meta-Analysis		
			OR	95% CI	P-value	OR	95% CI	P-value	OR	95% CI	P-value
LATE-NC	rs6564590	Add.	2.09	1.50-2.94	8.65×10^{-6}	1.28	1.07-1.54	0.00783	1.43	1.22-1.68	1.07×10^{-5}
	rs7404901	Add.	1.94	1.38-2.74	0.00011	1.31	1.09-1.58	0.00435	1.44	1.22-1.69	1.56×10^{-5}
HS	rs9925100	Rec.	2.32	1.19-4.35	0.01481	2.49	1.48-4.07	0.00087	2.44	1.63-3.61	1.34×10^{-5}
	rs9930659	Rec.	2.87	1.54-5.20	0.00112	1.98	1.20-3.17	0.00829	2.29	1.57-3.34	1.82×10^{-5}
B-ASC	rs4435266	Dom.	0.75	0.64-0.88	0.00041	0.71	0.54-0.94	0.01447	0.74	0.64-0.85	2.02×10^{-5}

NACC = National Alzheimer's Coordinating Center; ROSMAP = Religious Orders Study and Rush Memory and Aging Project; HS = hippocampal sclerosis; LATE-NC = limbic-predominant age-related TDP-43 encephalopathy neuropathological changes; B-ASC = brain arteriolosclerosis; MOI = mode of inheritance; Rec. = recessive MOI; Add. = additive MOI; Dom. = dominant MOI; OR = odds ratio; and CI = confidence interval.

Figure 2.1. LocusZoom plots of the WWOX/MAF region ± 250 kb for A neuritic plaques and B neurofibrillary tangles, both assuming an additive MOI. Meta-analytic variant-level p-values were adjusted for age at death, sex, Alzheimer's Disease Genetics Consortium (ADGC) cohort or Religious Orders Study and Memory and Aging Project (ROSMAP) study, and first three genetic principal components and meta-analyzed across the NACC and ROSMAP cohorts. The horizontal line at 4.44 represents the Bonferroni-corrected threshold for significance for the WWOX/MAF locus ± 250 kb. The blue region on the gene window highlights the location of rs55751884, the variant previously found to be genome-wide suggestive for HS; the green region on the gene window highlights the location of rs62039712, the variant previously found to be genome-wide significant for clinical AD; and the red region on the gene window highlights the location of rs450674, an MAF variant recently found to be associated with clinical AD.

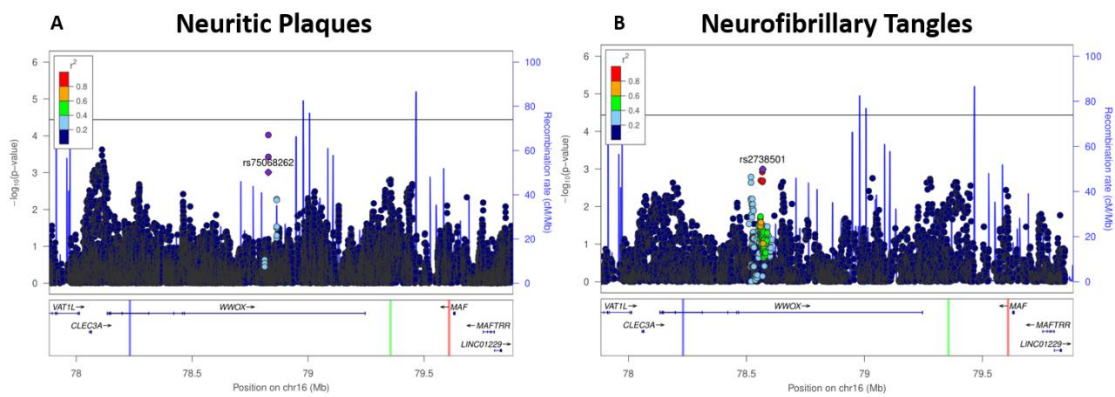
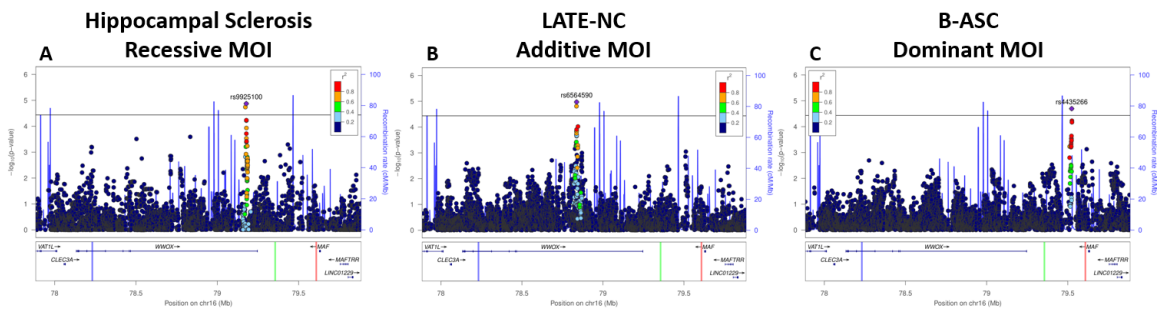


Figure 2.2. LocusZoom plots of the WWOX/MAF region +/- 250kb for A hippocampal sclerosis (HS) assuming a recessive mode of inheritance (MOI), B limbic-predominant age-related TDP-43 encephalopathy neuropathological changes (LATE-NC) assuming an additive MOI, and C brain arteriolosclerosis assuming a dominant MOI. Variant-level p-values were adjusted for age at death, sex, Alzheimer's Disease Genetics Consortium (ADGC) cohort or Religious Orders Study and Memory and Aging Project (ROSMAP) study, and first three genetic principal components. The horizontal line at 4.44 represents the Bonferroni-corrected threshold for significance for the WWOX/MAF locus \pm 250kb. The blue region on the gene window highlights the location of rs55751884, the variant previously found to be genome-wide suggestive for HS; the green region on the gene window highlights the location of rs62039712, the variant previously found to be genome-wide significant for clinical AD; and the red region on the gene window highlights the location of rs450674, an MAF variant recently found to be associated with clinical AD.



CHAPTER 3. ANALYSIS OF GENES (*TMEM106B*, *GRN*, *ABCC9*, *KCNMB2*, AND *APOE*)
IMPLICATED IN RISK FOR LATE-NC AND HIPPOCAMPAL SCLEROSIS PROVIDES
PATHOGENETIC INSIGHTS

3.1 Abstract

Transactive response DNA binding protein 43kDa (TDP-43) proteinopathy is commonly found in aged brains at autopsy and is associated with cognitive impairment. Limbic-predominant age-related TDP-43 encephalopathy neuropathologic change (LATE-NC) is the most prevalent subtype of TDP-43 proteinopathy, affecting ~1/3rd of aged persons. LATE-NC often co-occurs with hippocampal sclerosis (HS) pathology. It is currently unknown why some individuals with LATE-NC develop HS while others do not, but genetics may play a role. Previous studies found associations between LATE-NC phenotypes and specific genes: *TMEM106B*, *GRN*, *ABCC9*, *KCNMB2*, and *APOE*. Data from research participants with genomic and autopsy measures from the National Alzheimer's Coordinating Center (NACC; total n=631 subjects included) and the Religious Orders Study and Memory and the Rush Aging Project (ROSMAP; total n=780 included) were analyzed in the current study. Our goals were to reevaluate disease-associated genetic variants using newly collected data and to query whether the specific genotype/phenotype associations could provide new insights into disease-driving pathways. Research subjects included in prior LATE/HS genome-wide association studies (GWAS) were excluded. Single nucleotide variants (SNVs) within 10kb of *TMEM106B*, *GRN*, *ABCC9*, *KCNMB2*, and *APOE* were tested for association with HS and LATE-NC. Significantly associated SNVs were identified. When results were meta-analyzed, *TMEM106B*, *GRN*, and *APOE* had significant gene-based associations with both LATE and HS, whereas *ABCC9* had significant associations with HS only. In a sensitivity analysis limited to LATE-NC+ cases,

ABCC9 variants were again associated with HS. By contrast, the associations of TMEM106B, GRN, and APOE with HS were attenuated when adjusting for TDP-43 proteinopathy, indicating that these genes may be associated primarily with TDP-43 proteinopathy. In summary, using data not included in prior studies of LATE or HS genomics, we replicated several previously reported gene-based associations and found novel evidence that specific risk alleles can differentially affect LATE-NC and HS.

3.2 Introduction

The present study focused on genetic contributions to transactive response DNA binding protein 43kDa (TDP-43) proteinopathy and hippocampal sclerosis (HS). One or both of these pathologic features are observed in ~30% of brains among persons >80 years at death [9]. The TDP-43 protein serves multiple functions in gene expression regulation at the levels of both transcription and translation [9, 81-83]. TDP-43 proteinopathy (aberrantly misfolded and mislocalized TDP-43 protein) is strongly associated with cognitive impairment [11, 13]. This pathologic hallmark was discovered in diseases that are now considered to be a clinical-pathologic spectrum that includes amyotrophic lateral sclerosis (ALS) and frontotemporal lobe degeneration with TDP-43 (FTLD-TDP) [84].

HS is a pathologic finding characterized by selective neuronal loss and gliosis of the hippocampal formation [16, 17]. First described in epilepsy, HS is a descriptive and relatively nonspecific term used in both neuropathologic and neuroradiographic practice. However, in a subset of cases with HS, TDP-43 proteinopathy is also present [11, 16, 17, 85].

Limbic-predominant age-related TDP-43 encephalopathy (LATE) is a highly prevalent disease entity characterized by TDP-43 proteinopathy and cognitive impairment in aged populations [9]. LATE is not a subtype of FTLN-TDP because the associated disease(s) is not the frontotemporal dementia (FTD) clinical syndrome; rather, the presence of the neuropathologic changes underlying LATE (LATE-NC) is an amnesic dementia syndrome [9, 13-15]. HS pathology commonly co-occurs with LATE-NC and was the first neuropathologic change associated with the condition [86, 87]. In one autopsy cohort, the odds of TDP-43 proteinopathy was >80 times higher in HS cases relative to controls [16]. However, some persons with LATE-NC have no HS, segmental/patchy HS, or unilateral HS [88]. It is currently unknown why some individuals with LATE-NC develop HS pathology while others do not, but genetics may contribute to the pathogenesis.

Several genes and single nucleotide variants (SNVs) have been linked with LATE-NC phenotypes [9]. Risk for HS was previously associated with SNVs that are also known FTLN-TDP risk alleles, including rs5848 from the *GRN* gene on chromosome 17 and rs1990622 near the *TMEM106B* gene on chromosome 7 [25, 33-36, 89]. In a genome-wide association study (GWAS), a SNV in the *ABCC9* gene (rs704178/rs704180) on chromosome 12 was associated with HS risk [28]. A separate GWAS found that rs9637454, an SNV in the *KCNMB2* gene on chromosome 3, was associated with HS risk [37]. Additional evidence exists linking the *APOE* ϵ 4 allele, a strong risk factor for Alzheimer's disease (AD), with increased HS and LATE-NC risk [32, 38, 41]. A study analyzing gene-based associations between the *GRN*, *TMEM106B*, *ABCC9*, and *KCNMB2* genes and HS found Bonferroni-corrected significant associations for *ABCC9* assuming a recessive mode of inheritance (MOI) and nominally significant associations with *GRN*, *TMEM106B*, and

KCNMB2 [27]. However, a separate study replicated the associations between *GRN* and *TMEM106B* SNVs with LATE-NC, but did not find an association between an *ABCC9* variant and LATE-NC or HS pathologies [26]. To the best of our knowledge, there has not been a prior study that found genomic associations with LATE-NC but not HS or vice versa.

In the current study, we analyzed genomic data from the Alzheimer’s Disease Genetics Consortium (ADGC) along with clinical and pathological data from the National Alzheimer’s Coordinating Center (NACC) and the Rush University Religious Orders Study and Memory and Aging Project (ROSMAP) to investigate the associations between prior identified putative risk genes – *KCNMB2*, *TMEM106B*, *ABCC9*, *GRN*, and *APOE* – and LATE-NC. By only analyzing participants not included in our prior studies [27, 28], we sought to test whether or not previously reported LATE-NC risk genes can be replicated for LATE-NC neuropathologic phenotypes (specifically, TDP-43 proteinopathy and HS) while also identifying novel risk SNVs.

3.3 Material and Methods

3.3.1 Study Participants

Representative photomicrographs were taken, showing results from research participants with LATE-NC and LATE-NC+HS, in the University of Kentucky AD Research Center Autopsy cohort, using methods as previously described [29].

Phenotypic data from NACC (March 2021 data freeze) were linked with genotype data from the ADGC. Individuals who died at age 65 years or older were included. Similar to other studies using NACC data [58], individuals were excluded from the NACC cohort if

at least one of 19 rare brain diseases were diagnosed (**Supplemental Table 3.1**) or if they were missing any adjustment variables or both endophenotypes under study.

The ROSMAP study has been described in detail elsewhere [59]. Briefly, data were acquired from two well-characterized cohort studies of aging and dementia. The Religious Orders Study (ROS), begun in 1994, and the Rush Memory and Aging Project (MAP), begun in 1997, involve older adults who enrolled without dementia, agreed to annual clinical evaluations and organ donation at death, and signed an Anatomical Gift Act for brain donation. Written informed consent was obtained from participants, and research was carried out in accordance with Institutional Review Board (IRB)-approved protocols. ROSMAP data are available online at the Rush Alzheimer's Disease Center Resource Sharing Hub (<https://www.radc.rush.edu/>), as well as on the Accelerating Medicines Partnership-Alzheimer's Disease (AMP-AD) Knowledge Portal (syn3219045).

For both the NACC and ROSMAP datasets, individuals were excluded from the analyses if they were included in either of two previous studies of HS genomics [27, 28]. In ROSMAP, participants were excluded based on IID if they were included in the Nelson et al. HS GWAS from 2014. In NACC, HS and TDP-43 were defined using variables from the v10 NACC Neuropathology (NP) dataset which were not available for the participants included in the previous studies. Thus, the NACC and ROSMAP participants included in the current study are a true replication cohort for these earlier HS genomics studies.

3.3.2 Neuropathological Endophenotype Definitions

In the NACC NP dataset, LATE-NC was defined as either present or absent using the “distribution of TDP-43 immunoreactive inclusions” variables indicating if TDP-43

proteinopathy was observed in either the hippocampus (NPTDPC NACC field), entorhinal/inferior temporal cortex (NPTDPD), or neocortex (NPTDPE) in a case lacking overall diagnosis of FTLT-DTP. A LATE-NC case was defined as definitely having TDP-43 in the hippocampus, entorhinal/inferior temporal cortex, or neocortex. LATE-NC was considered unknown if TDP-43 data were unavailable in all three regions. HS was defined as either present or absent based on the “hippocampal sclerosis of CA1 and/or subiculum” (NPHIPSCL) variable using the “unilateral,” “bilateral,” and “present but laterality not assessed” response categories.

In the ROSMAP data set, LATE-NC was defined dichotomously using the “TDP-43 stage” (tdp_st4) variable and collapsing the 2nd and 3rd stages in cases lacking FTLT-DTP. HS was defined dichotomously by the “hippocampal sclerosis was rated as definitely present with CA1 region affected” response category of the “definite presence of typical hippocampal sclerosis” (hspath_typ) variable.

3.3.3 Quality Control of Genotype Data

For NACC participants, genomic data from the ADGC imputed using the Haplotype Reference Consortium (ADGC-HRC) were used [60]. The genetic data for ROSMAP were also imputed using the HRC and the methods have been described in detail elsewhere [61]. Standard GWAS quality control (QC) procedures were performed separately on the ADGC and ROSMAP genotype data using PLINK1.9 [62, 63]. SNVs were excluded if they were missing in more than 5% of samples, if they had a minor allele frequency less than 1%, or if they had Hardy-Weinberg Equilibrium (HWE) p-values $< 1 \times 10^{-6}$ among AD controls. Individuals were excluded if they were missing more than 5%

of genotypes. Two individuals were considered related if they had an identity by descent measure of at least 0.25, which indicates that they are second-degree relatives. For related pairs, the individual with the lowest call rate was excluded.

NACC and ROSMAP genotype data were separately merged with 1000 Genomes Project Phase 3 data. Principal components (PCs) were calculated for the merged data sets using the “pca” procedure in PLINK1.9, and the first two PCs were plotted. The ADGC-HRC and ROSMAP individuals with first and second PCs that overlapped with those of the 1000 Genomes individuals of known European ancestry were identified and all other individuals were excluded from the analysis.

3.3.4 Variant-Level Associations

All statistical analyses were conducted in R programming language [64], version 4.0.4. Associations between each endophenotype and each SNV were conducted separately in the NACC and ROSMAP datasets using binary logistic regression models assuming each of the three most common MOI: additive, dominant, and recessive. SNVs were excluded from the analyses if they were multiallelic or if there were fewer than 15 minor alleles present across all participants. All regression models were fit using the glm function in R assuming a binomial distribution and a logit link function and were adjusted for age at death, sex, ADGC data selection round (for NACC data) or ROS/MAP study (for ROSMAP data), and the first three genetic PCs. Odds ratios (OR) were calculated for each SNV. Since some endophenotypes were only available in a subset of participants, PCs were calculated separately for each endophenotype. NACC and ROSMAP SNV-level results were meta-analyzed using a fixed-effect, inverse-variance meta-analysis via the metagen

function from the meta R package, version 4.18-0 [65]. For targeted analyses of previously reported SNVs, an additive MOI was assumed unless there existed previous evidence of association with another MOI. Additionally, LATE-NC-by-SNV interaction terms were tests for models of HS and were removed if they failed to reach statistical significance ($p < 0.05$). Plots of cohort-specific and meta-analyzed SNV-level p-values were created using LocusZoom Standalone, version 1.4 (https://genome.sph.umich.edu/wiki/LocusZoom_Standalone) [66], and the ggplot2 R package, version 3.3.3 [90]. Linkage disequilibrium estimates were computed using LDlink with the CEU population (<https://ldlink.nci.nih.gov/>) [67]. Jaccard similarity coefficients were used to estimate the similarity between binary variables and were calculated by dividing the size of their intersection by the size of their union via the clusteval R package, version 0.1 [91].

3.3.5 Gene-Based Associations

Gene boundaries for *KCNMB2*, *TMEM106B*, *ABCC9*, *GRN*, and *APOE* were defined based on their canonical transcripts using the Genome Reference Consortium Human Build 37 (GRCh37/hg19) gene range list from PLINK (<https://www.cog-genomics.org/plink/1.9/resources>). All genes were flanked by an additional 10kb to include potential regulatory regions. See **Supplemental Table 3.2** for the positions used to define the gene boundaries.

For each gene, endophenotype, and MOI, all SNV-level p-values were combined using the aggregated Cauchy association test (ACAT) [92]. All ACAT analyses were run using R functions provided by the authors (<https://github.com/yaowuliu/ACAT>). Equal

weights were assumed for all SNVs in the ACAT analyses and statistical significance was defined as a p-value < 0.05 .

3.3.6 SNV Prioritization and Follow-Up Analyses

Prioritized SNVs were identified using a Bonferroni-corrected threshold for significance that accounts for the effective number of independent tests in a given genetic region. The effective number of independent tests in a region was calculated for each endophenotype using the method of Gao et al. [68]. Briefly, Pearson's correlation coefficient was calculated for all pairs of SNVs and these coefficients were placed in a square matrix. The eigenvalues of the matrix were then computed and ordered from largest to smallest and the effective number of independent tests was defined to be the smallest number of ordered eigenvalues that account for 99.5% of the sum of all eigenvalues. The Bonferroni-corrected threshold for identifying prioritized SNVs in a given genetic region was defined as 0.05 divided by the largest estimated number of independent tests in the region.

Prioritized SNVs were investigated for expression quantitative trait loci (eQTL) associations using the Genotype-Tissue Expression (GTEx) Project's V8 public data [69], the BRAINEAC Brain eQTL Almanac (<http://braineac.org/>) [70], and Functional Annotation of Human Long Noncoding RNAs via Molecular Mapping (FANTOM5) database (data accessed via: <https://www.ebi.ac.uk/gxa/experiments/E-MTAB-3358/Results>). Prioritized SNVs were also investigated for associations with other molecular mechanisms using the INFERRing the molecular mechanisms of NONcoding

genetic variants (INFERNO) software assuming a threshold on r^2 of 0.5 and a threshold on LD block size of 500kb (<http://inferno.lisanwanglab.org/index.php>) [71].

3.3.7 Sensitivity Analyses

The dependency of the study's results on several analytic choices were investigated. All gene-based analyses were also conducted assuming 0kb and 25kb of flanking around each gene. All *APOE* SNV analyses were also adjusted for the number of *APOE* $\epsilon 4$ alleles to determine if associations were independent of $\epsilon 4$ risk.

3.4 Results

The phenotypes of interest in the current study are autopsy-confirmed LATE-NC and HS. Specific examples of those pathologies are depicted in **Figure 3.1**. Some brains have LATE-NC without HS (**Figure 3.1B**). However, individuals with LATE-NC are at increased risk of having comorbid HS (**Figure 3.1C**).

In the ROSMAP data set, a total of $n=795$ individuals had available data for at least one of the endophenotypes along with GWAS data and were not included in earlier studies of HS [27, 28]. In the NACC data set, $n=633$ individuals had available data for at least one of the endophenotypes along with GWAS data and were not included in the earlier studies of HS [27, 28]. **Table 3.1** shows a summary of individual characteristics and endophenotypes for both NACC and ROSMAP participants. ROSMAP participants tended to be older at death ($p<0.001$), were more likely to be female ($p<0.001$), and were less likely to be an HS case ($p=0.007$) than NACC participants. HS was less prevalent than LATE-NC in both cohorts (NACC: HS 14.1%, LATE-NC 29.4%; ROSMAP: HS 9.4%,

LATE-NC 33.2%). HS and LATE-NC cases in ROSMAP tended to be older at death (both $p < 0.001$) and were less likely to be male ($p = 0.054$ and $p < 0.001$, respectively) than controls. There were no identified statistically significant differences in basic demographic characteristics between HS and/or LATE-NC cases and controls in NACC.

Persons with HS tended to also have LATE-NC and the reverse was also true among individuals in both datasets (Jaccard coefficients of 0.589 and 0.575 in NACC and ROSMAP, respectively); see **Figure 3.2**. Of the 732 ROSMAP participants with available case data for both LATE-NC and HS, 93% of HS cases were also LATE-NC cases. Of the 410 NACC participants with available case data for both LATE-NC and HS, 73% of HS cases were also LATE-NC cases.

Across the *KCNMB2*, *TMEM106B*, *ABCC9*, *GRN*, and *APOE* genes, each flanked by 10kb, a total of 1,580 SNVs passed QC in NACC while 1,532 SNVs passed QC in ROSMAP. A total of 1,438 SNVs were shared between NACC and ROSMAP and were included in the meta-analysis (**Supplemental Table 3.2**).

3.4.1 Gene-Based Associations

The adjusted meta-analyzed, SNV-level results were combined within genes via ACAT to obtain gene-based p-values. At the gene level, *TMEM106B* and *APOE* were significantly associated with both HS and LATE-NC while *ABCC9* and *GRN* were significantly associated with HS only (**Table 3.2**). Neither HS nor LATE-NC were significantly associated with *KCNMB2*. The meta-analyzed gene-based results were largely similar to when they were conducted separately in the NACC and ROSMAP datasets.

Additionally, these results were largely unchanged when 0kb and 25kb of flanking were added to each gene.

3.4.2 Prioritized SNVs and Follow-Up Analyses

The effective number of independent tests for *TMEM106B* \pm 10kb was estimated to be 25, *GRN* \pm 10kb was estimated to be 16, *KCNMB2* \pm 10kb was estimated to be 104, *APOE* \pm 10kb was estimated to be 14, and *ABCC9* \pm 10kb was estimated to be 71. The Bonferroni-corrected thresholds for a genetic region was calculated by dividing 0.05 by the corresponding estimated effective number of independent tests in the region.

One hundred and ten SNVs in the *TMEM106B* \pm 10kb locus had adjusted meta-analytic associations with HS or LATE-NC less than the Bonferroni-corrected threshold (**Figure 3.3A**). At the *TMEM106B* \pm 10kb locus, rs7781670 had the smallest adjusted meta-analytic p-value for LATE-NC assuming an additive MOI ($p=2.97 \times 10^{-5}$). rs7781670 also met the Bonferroni-corrected threshold for the *TMEM106B* \pm 10kb locus for HS when assuming a recessive MOI ($p=1.63 \times 10^{-3}$) and was a significant eQTL in GTEx for *TMEM106B* in the cerebellum ($p=4.7 \times 10^{-7}$) and the cortex ($p=2.6 \times 10^{-5}$). In INFERNO, these prioritized SNVs were associated with both eQTLs and Roadmap enhancers in blood, connective, and epithelial tissues and just Roadmap enhancers in brain, heart, immune organ, liver, and skeletal tissues, among others.

Fourteen SNVs in the *GRN* \pm 10kb locus had adjusted meta-analytic associations with HS or LATE-NC less than the Bonferroni-corrected threshold (**Figure 3.3B**). rs5848 had the smallest adjusted meta-analytic p-value in the *GRN* \pm 10kb locus and met the Bonferroni-corrected threshold for HS (additive MOI $p=2.16 \times 10^{-4}$; recessive MOI

$p=1.91 \times 10^{-4}$). rs5848 also had the smallest adjusted meta-analytic p-value for LATE-NC in the *GRN* \pm 10kb locus, but it did not meet the Bonferroni-corrected threshold. In GTEx, rs5848 was a significant eQTL for GRN expression in numerous tissues including thyroid ($p=2.2 \times 10^{-16}$), caudate ($p=2.0 \times 10^{-12}$), cortex ($p=2.0 \times 10^{-9}$), and frontal cortex ($p=4.4 \times 10^{-9}$). In INFERNO, these prioritized SNVs were associated with both eQTLs and Roadmap enhancers in adipose, connective, endocrine, heart, and nervous tissues, with just eQTLs in blood vessel tissue, and with just Roadmap enhancers in brain, blood, immune organ, liver, and skeletal muscle tissues, among others.

No SNVs in the *KCNMB2* \pm 10kb locus had adjusted meta-analytic associations with HS or LATE-NC that met the Bonferroni-corrected threshold (**Supplemental Figure 3.1**).

The *APOE* \pm 10kb locus was strongly associated with LATE-NC. Four SNVs (rs429358, rs769449, rs10414043, and rs7256200), all in high linkage disequilibrium with one another (all $r^2 > 0.95$), had adjusted meta-analytic associations with LATE-NC that met the Bonferroni-corrected threshold assuming an additive MOI (all p-values $\leq 2.56 \times 10^{-8}$) (**Figure 3.4**). While none of the *APOE* SNVs were associated with *APOE* expression levels in the evaluated data sets, rs769449 and rs10414043 were significant sQTLs in GTEx for *TOMM40* in cerebellar hemisphere tissue ($p=4.0 \times 10^{-10}$ and $p=1.4 \times 10^{-5}$, respectively). In INFERNO, these prioritized SNVs were associated with both Roadmap and FANTOM5 enhancers in adipose, blood, brain, connective, epithelial, liver, nervous, skeletal muscle, smooth muscle, and stem cell tissues and with just Roadmap enhancers in endocrine, heart, and immune organ tissues, among others.

The *ABCC9* ± 10kb locus was most strongly associated with HS and contained 13 SNVs with adjusted meta-analytic p-values for HS less than the Bonferroni-corrected threshold (**Figure 3.5**). rs1914361 had the smallest adjusted meta-analytic p-value with HS assuming a recessive MOI ($p=1.70 \times 10^{-4}$). In prior studies with cohorts of research subjects that did not overlap with the current study, the *ABCC9*/HS association was strongest for the recessive MOI models [27, 28, 31]. All other SNVs that also met the Bonferroni-corrected threshold when assuming a recessive MOI were in high linkage disequilibrium with rs1914361 (all $r^2 > 0.75$). rs1914361 was a significant eQTL in the GTEx data set for the expression of *ABCC9* in several tissues, including brain (nucleus accumbens, caudate, cortex, and putamen) and artery tissues (tibial and aorta) (**Figure 3.6A**). Notably, rs1914361 minor alleles were positively correlated with *ABCC9* expression in brain tissues (**Figure 3.6B**) and negatively correlated with *ABCC9* expression in artery tissues (**Figure 3.6C**). Furthermore, relative to rs704178, a previously identified *ABCC9* HS SNV, rs1914361 had a similarly strong association with *ABCC9* gene expression in GTEx (rs704178: $p=4.00 \times 10^{-13}$; rs1914361: $p=7.10 \times 10^{-12}$) and a stronger association with *ABCC9* gene expression in BRAINEAC (rs704178: $p=6.80 \times 10^{-4}$; rs1914361: $p=2.10 \times 10^{-7}$) (**Table 3.3**). In INFERNO, these prioritized SNVs were associated with Roadmap enhancers in adipose, blood vessel, connective, heart, liver, skeletal muscle, and smooth muscle tissues, among others.

3.4.3 SNV-Level Regression Analyses

In their respective regression models, the *GRN* SNV rs5848 ($p=0.010$), the *APOE* SNV rs769449 ($p<0.001$), and *APOE* ε4 carrier status ($p<0.001$) all had nominally

significant adjusted meta-analytic associations with LATE-NC and the *TMEM106B* SNV rs7781670 had a borderline-significant adjusted meta-analytic association with LATE-NC ($p=0.057$) (**Table 3.4**). All odds ratio estimates were consistent across NACC and ROSMAP with the exception of the *ABCC9* SNV rs1914361 when assuming a recessive MOI (NACC: OR=0.98; ROSMAP: OR=1.40). Notably, the odds ratio estimates for LATE-NC were very similar between the *APOE* SNV rs769449 (meta-analytic OR=1.95) and *APOE* $\epsilon 4$ carrier status (meta-analytic OR=2.05), which likely reflects the fact that rs769449 minor allele counts are strongly correlated with *APOE* $\epsilon 4$ counts (NACC: $r^2=0.746$; ROSMAP: $r^2=0.712$).

No LATE-NC-by-SNV interactions were significant in the adjusted HS models, so the interaction terms were removed. The *TMEM106B* SNVs (rs1990622 and rs7781670), the *GRN* SNV (rs5848), one of the *ABCC9* SNVs (rs1914361), the *APOE* SNV (rs769449), and *APOE* $\epsilon 4$ carrier status all had nominally significant, adjusted meta-analytic associations with HS (**Table 3.5**). When these models were adjusted for LATE-NC, all models had nominally significant adjusted meta-analytic associations with HS with the exception of the *APOE* SNV (rs769449) and *APOE* $\epsilon 4$ carrier status (**Table 3.5, Figure 3.7**), suggesting that the association between *APOE* status and HS is related to a more direct interaction between *APOE* and LATE-NC (i.e., TDP-43 proteinopathy). Additionally, the association between HS and the *ABCC9* SNV rs704178 becomes nominally significant with larger odds ratio estimates when adjusted for LATE-NC.

An issue raised by the *ABCC9*/HS association results was whether this correlation was driven by cases lacking LATE-NC, i.e. the minority of cases with HS pathology that lacked TDP-43 proteinopathy. A separate sensitivity analysis was performed that excluded

just the cases with HS pathology that lacked LATE-NC. Results are shown in **Supplemental Table 3.3**, which may be compared with **Table 3.5**. The odds ratio estimates for the association between *ABCC9* risk variants and HS pathology was essentially unchanged by removing the LATE-NC-HS+ cases.

3.5 Discussion

Using large genetic data sets and autopsy-derived data, we demonstrated that the neuropathological endophenotypes of LATE-NC and HS showed replication for associations between *TMEM106B*, *GRN*, *APOE*, and/or *ABCC9*. Interestingly, *ABCC9* was not associated with LATE-NC but was with HS pathology. Our study adds to the growing body of literature on the overlapping genetics of HS and LATE-NC while also highlighting several genetic loci unique to each disease entity.

We replicated significant gene-based associations between HS and the *TMEM106B*, *ABCC9*, *GRN*, and *APOE* genes along with the rs7781670 (*TMEM106B*) and rs5848 (*GRN*) SNVs. Furthermore, we identified novel SNV-level associations between LATE-NC and rs7781670 and rs769449. The association of LATE-NC and rs7781670 is intriguing since it was also recently associated with clinical AD in a large AD GWAS [54]. We found no evidence to support the hypothesis that *KCNMB2* is a risk gene for either LATE-NC or HS pathologies. However, we note that the sample size of the present study was suitable to detect only relatively large genotype/phenotype associations.

There is an emerging consensus that mixed pathologies are highly prevalent in elderly populations, and there are complex relationships between genotypes and phenotypes. The finding that variants in *GRN*, *TMEM106B*, and *APOE* genes are associated

with several neuropathological endophenotypes fits in with recent studies looking at genetic pleiotropy in neurological conditions [39]. Pleiotropic effects have been observed among AD-related neuropathological changes like neuritic plaques, neurofibrillary tangles, and cerebral amyloid angiopathy [76] as well as between LATE-NC and FTLT-TDP [9].

Since the large majority of HS cases were also LATE-NC cases in the current study (**Figure 3.2**), it was striking that some risk genes and SNVs were found to only be associated with HS and not LATE-NC—and vice versa, when statistical models were applied. We did identify several genes that are associated with both neuropathologic endophenotypes. Specifically, the *TMEM106B*, *GRN*, and *APOE* SNVs appear to predispose individuals to LATE-NC (**Figure 3.9A**). Our data indicate that the associations between HS and SNVs in the *TMEM106B*, *GRN*, and *ABCC9* genes remain statistically significant in a model that adjusts for the presence of LATE-NC (**Table 3.5**). However, the impact of *TMEM106B* and *GRN* on HS appeared to be attenuated in a statistical model that included TDP-43 proteinopathy, suggesting that their impact on HS may be mediated by their role in LATE-NC. How these genetic SNVs can impact HS secondarily or independently of LATE-NC is not currently known.

While several *ABCC9* SNVs have been found to be associated with HS, including rs704178 and rs704180, this is the first study to report an association between the *ABCC9* SNV rs1914361 and HS. Notably, rs1914361 was found to be associated with HS in two of the three included cohorts of the original HS GWAS [28], but it was not included in the downstream analyses since its association with HS wasn't nominally significant in all three cohorts (data not published). It is important to note that prior study involved a completely different set of study participants but the “direction” of the effect in all cohorts studied was

the same. Since rs1914361 was found to also be significantly associated with the expression of *ABCC9* (**Table 3.3**) and is not in strong linkage disequilibrium with rs704178 ($r^2=0.176$), the two loci may represent independent *ABCC9* HS risk SNVs.

We also identified divergent patterns in the tissue-level gene expressions of *ABCC9* and its homologous gene, *ABCC8*. Similar results were observable in both the GTEx and FANTOM5 databases (**Figure 3.8**). While *ABCC9* appears to be highly expressed in vascular and smooth muscle tissues (**Figure 3.6A**) and modestly expressed in CNS tissues, the paralogous *ABCC8* gene tends to be more highly expressed in CNS tissue and less abundantly expressed in vascular and smooth tissues. The proteins encoded by these genes both serve to regulate the “KATP” potassium channels [93], but their divergent expression patterns suggest that *ABCC9*-related disease risk may act via vascular pathogenetic mechanisms [30].

The current study adds to a growing body of literature suggesting that LATE-NC is a potential precursor to HS [9]. It is yet to be seen how exactly the *APOE* gene and AD-type changes interact with other pathologies, but one hypothesis is that *APOE* and AD predispose an individual to LATE-NC, which then drives an individual towards severe LATE-NC and HS (**Figure 3.9B**). It has been found that TDP-43 proteinopathy localizes to tangle-like structures in many cases with ADNC [40]. Further autopsy-based studies with larger sample sizes are needed. We note that studies that focus on the LATE-NC phenotype with HS should optimally include relatively large numbers of individuals who died past 85 years of age.

There are both limitations and strengths to the present study. Because of the characteristics of the sample (largely Caucasian, drawn from a number of different research

centers), the degree to which findings are generalizable is unknown, especially with respect to individuals of other ancestries. While this work aims to replicate previous associations, there are many models considered which can inflate false positive rates. Additionally, it is difficult to know if the associations identified in the current study are independent of ADNC, especially the significant associations between *APOE* and LATE-NC. However, this association likely still highlights the strong associations that exist between AD and other neurodegenerative diseases, which is interesting in itself. We also note that all the included subjects had high-quality neuropathologic workup for TDP-43 proteinopathy and HS, and all the ADGC subjects were autopsied during 2014 and later. These study design elements constitute strengths of the current study.

Table 3.1. Participant characteristics stratified by hippocampal sclerosis (HS) and limbic-predominant age-related TDP-43 encephalopathy neuropathological changes (LATE-NC) case status.

	NACC			ROSMAP		
	Number of Participants (%)	Age at Death, Mean (SD)	Female, N (%)	Number of Participants (%)	Age at Death, Mean (SD)	Female, N (%)
HS						
Overall	N=631	85.9 (8.3)	319 (50.6)	N=780	88.7 (7.2)	525 (67.3)
No	542 (85.9)	85.9 (8.4)	270 (49.8)	707 (90.6)	88.3 (7.2)	468 (66.2)
Yes	89 (14.1)	86.0 (7.5)	49 (55.1)	73 (9.4)	92.0 (6.4)	57 (78.1)
LATE-NC						
Overall	N=512	85.1 (7.9)	207 (50.2)	N=747	89.1 (7.1)	506 (67.7)
No	291 (70.6)	84.9 (8.1)	138 (47.4)	499 (66.8)	87.9 (7.3)	315 (63.1)
Yes	121 (29.4)	85.4 (7.3)	66 (57.0)	248 (33.2)	91.5 (6.1)	191 (77.0)

NACC = National Alzheimer's Coordinating Center; ROSMAP = Religious Orders Study and Rush Memory and Aging Project; SD = standard deviation; HS = hippocampal sclerosis; LATE-NC = limbic-predominant age-related TDP-43 encephalopathy neuropathological changes.

Table 3.2. Aggregated Cauchy association test (ACAT) gene-based p-values for hippocampal sclerosis (HS) and limbic-predominant age-related TDP-43 encephalopathy neuropathological changes (LATE-NC). Each gene is flanked by 10kb. All SNV-level analyses were adjusted for sex, age at death, cohort/study, and the first three genetic principal components and meta-analyzed across National Alzheimer's Coordinating Center (NACC) and Religious Orders Study and Rush Memory and Aging Project (ROSMAP) participants.

Chr.	Gene	Endophenotype	MOI		
			Additive	Dominant	Recessive
3	<i>KCNMB2</i>	HS	0.718	0.632	0.478
		LATE-NC	0.980	0.995	0.473
7	<i>TMEM106B</i>	HS	0.006	0.052	0.005
		LATE-NC	<0.001	0.004	<0.001
12	<i>ABCC9</i>	HS	0.036	0.072	0.006
		LATE-NC	0.901	0.440	0.912
17	<i>GRN</i>	HS	0.004	0.348	0.003
		LATE-NC	0.164	0.628	0.069
19	<i>APOE</i>	HS	0.014	0.017	0.333
		LATE-NC	<0.001	<0.001	0.064

Chr. = chromosome; HS = hippocampal sclerosis; LATE-NC = limbic-predominant age-related TDP-43 encephalopathy neuropathological changes; MOI = mode of inheritance.

Table 3.3. Most significant expression quantitative trait loci (eQTL) p-values for *ABCC9* in BRAINEAC and GTEx databases.

Gene	SNV	Most Significant eQTL P-value	
		BRAINEAC	GTEx
<i>ABCC9</i>	rs704178	6.80E-04	4.00E-13
	rs1914361	2.10E-07	7.10E-12

eQTL = expression quantitative trait loci; GTEx = Genotype-Tissue Expression; SNV = single-nucleotide variant.

Table 3.4. Adjusted effects of single nucleotide variants (SNV) on limbic-predominant age-related TDP-43 encephalopathy neuropathological change (LATE-NC). All models adjust for sex, age at death, first three principal components and cohort/study. For rs1990622, rs7781670, and rs704178, the effect alleles are the risk-associated alleles and not the minor alleles.

Gene	MOI	SNV	Effect Allele	NACC		ROSMAP		Meta-Analysis		
				OR	P-value	OR	P-value	OR	95% CI	P-value
<i>TMEM106B</i>	Additive	rs1990622	A	1.39	0.051	1.08	0.484	1.16	(0.97, 1.39)	0.099
<i>TMEM106B</i>	Additive	rs7781670	C	1.47	0.024	1.09	0.415	1.19	(1.00, 1.43)	0.057
<i>GRN</i>	Additive	rs5848	T	1.40	0.042	1.23	0.089	1.29	(1.06, 1.56)	0.010
<i>ABCC9</i>	Additive	rs1914361	G	1.16	0.354	1.16	0.171	1.16	(0.97, 1.39)	0.098
<i>ABCC9</i>	Recessive	rs1914361	G	0.98	0.933	1.40	0.077	1.25	(0.92, 1.71)	0.151
<i>ABCC9</i>	Additive	rs704178	G	0.95	0.764	1.07	0.536	1.03	(0.86, 1.24)	0.732
<i>ABCC9</i>	Recessive	rs704178	G	0.80	0.433	1.16	0.394	1.05	(0.78, 1.40)	0.759
<i>APOE</i>	Additive	rs769449	A	1.70	0.004	2.22	<0.001	1.95	(1.51, 2.52)	<0.001
<i>APOE</i>	N/A	e4 Carrier	N/A	1.88	0.010	2.16	<0.001	2.05	(1.54, 2.74)	<0.001

NACC = National Alzheimer's Coordinating Center; ROSMAP = Religious Orders Study and Rush Memory and Aging Project; MOI = mode of inheritance; SNV = single-nucleotide variant; OR = odds ratio; CI = confidence interval.

Table 3.5. Adjusted effects of single nucleotide variants (SNV) on hippocampal sclerosis (HS). A separate regression model was fit for each variant, mode of inheritance (MOI), and limbic-predominant age-related TDP-43 encephalopathy neuropathological change (LATE-NC) adjustment. All models also adjust for sex, age at death, first three principal components and cohort/study. For rs1990622, rs7781670, and rs704178, the effect alleles are the risk-associated alleles and not the minor alleles.

Gene	MOI	SNV	Effect Allele	LATE-NC Adjusted	NACC		ROSMAP		Meta-Analysis		
					OR	P-value	OR	P-value	OR	95% CI	P-value
<i>TMEM106B</i>	Additive	rs1990622	A	Yes	1.23	0.396	1.61	0.017	1.44	(1.07, 1.95)	0.017
				No	1.43	0.044	1.55	0.019	1.49	(1.15, 1.91)	0.002
<i>TMEM106B</i>	Additive	rs7781670	C	Yes	1.20	0.456	1.53	0.029	1.39	(1.03, 1.88)	0.030
				No	1.47	0.034	1.50	0.028	1.48	(1.15, 1.91)	0.002
<i>GRN</i>	Additive	rs5848	T	Yes	1.37	0.168	1.37	0.123	1.37	(1.02, 1.84)	0.039
				No	1.67	0.004	1.43	0.057	1.56	(1.21, 2.00)	<0.001
<i>ABCC9</i>	Additive	rs1914361	G	Yes	1.92	0.005	1.31	0.152	1.52	(1.14, 2.03)	0.004
				No	1.64	0.004	1.35	0.092	1.49	(1.17, 1.90)	0.001
<i>ABCC9</i>	Recessive	rs1914361	G	Yes	3.87	<0.001	1.58	0.124	2.23	(1.42, 3.51)	<0.001
				No	2.69	<0.001	1.64	0.075	2.12	(1.45, 3.09)	<0.001
<i>ABCC9</i>	Additive	rs704178	G	Yes	1.51	0.079	1.52	0.034	1.52	(1.13, 2.04)	0.006
				No	1.09	0.618	1.42	0.059	1.23	(0.96, 1.57)	0.099
<i>ABCC9</i>	Recessive	rs704178	G	Yes	1.77	0.121	1.48	0.171	1.58	(1.02, 2.47)	0.042
				No	1.33	0.292	1.43	0.180	1.38	(0.95, 2.01)	0.090
<i>APOE</i>	Additive	rs769449	A	Yes	1.15	0.589	1.54	0.094	1.33	(0.93, 1.90)	0.118
				No	1.30	0.188	2.02	0.004	1.54	(1.14, 2.09)	0.005
<i>APOE</i>	N/A	e4 Carrier	N/A	Yes	1.74	0.114	1.34	0.318	1.49	(0.96, 2.31)	0.075
				No	1.79	0.024	1.91	0.018	1.84	(1.28, 2.66)	0.001

NACC = National Alzheimer's Coordinating Center; ROSMAP = Religious Orders Study and Rush Memory and Aging Project; MOI = mode of inheritance; SNV = single-nucleotide variant; LATE-NC = limbic-predominant age-related TDP-43 encephalopathy neuropathological change; OR = odds ratio; CI = confidence interval.

Figure 3.1. Photomicrographs of human hippocampi depict the main neuropathologic endophenotypes analyzed in the current study. Hippocampal sclerosis (HS) is evaluated with H&E stain (panels A,C,E), whereas LATE-NC is operationalized with phospho-TDP-43 immunohistochemistry (IHC; panels B, D, and F). All photomicrographs depict mid-level hippocampal sections dissected in the coronal plane. Panels A and B show stained brain sections from a woman (APOE e3/e4) who died at age 83; autopsy revealed neither LATE-NC nor HS. Panels C and D are from a man (APOE e3/e4) who died at age 93 with LATE-NC Stage 2. Panels E and F are from a woman (APOE e3/e3) who died at age 95 with LATE-NC Stage 2 and comorbid HS. Note the relatively atrophic hippocampal profile in Panel E in comparison to A or C (same scale bar); the HS+ profile in panel E also demonstrates parenchymal rarefaction which can be appreciated even at low magnification. Phospho-TDP-43 immunoreactive intraneuronal inclusions are highlighted with arrows in panels D and F. The representative photomicrographs were from research participants of the University of Kentucky AD Research Center. Scale bar = 2mm in A, C, and E, 75 microns in B, D, and F.

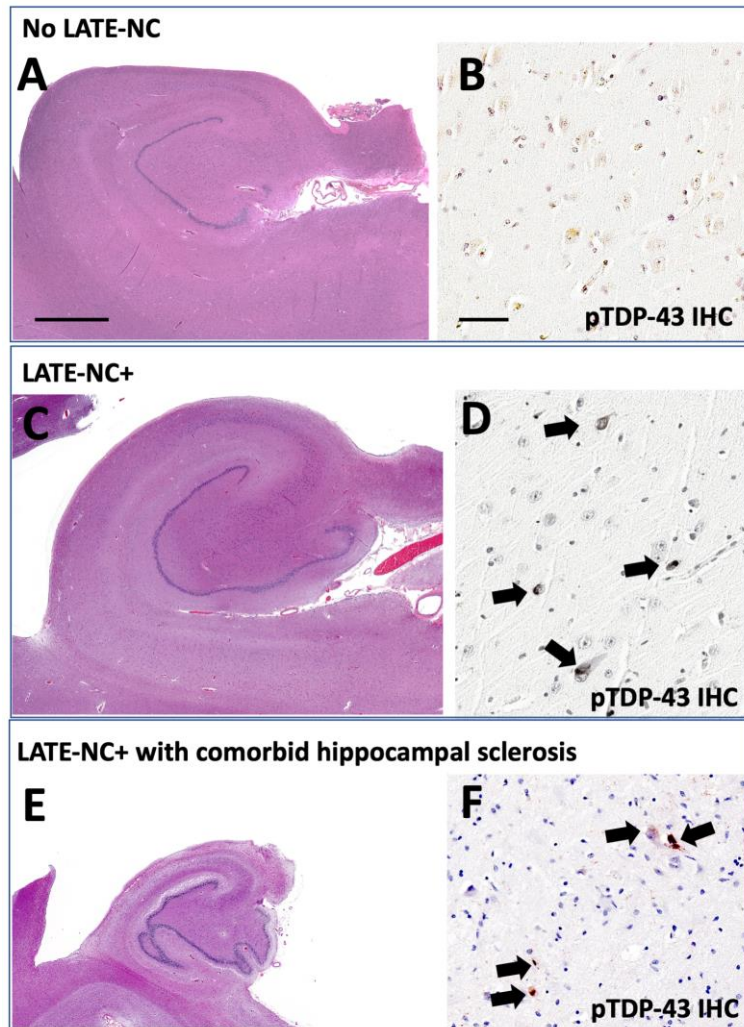


Figure 3.2. Venn diagrams of the overlap between limbic-predominant age-related TDP-43 encephalopathy neuropathological change (LATE-NC) and hippocampal sclerosis (HS) cases in **A.** National Alzheimer's Coordinating Center (NACC) and **B.** Religious Orders Study and Rush Memory and Aging Project (ROSMAP). Only participants with non-missing case data for both LATE-NC and HS are included. LATE-NC = limbic-predominant age-related TDP-43 encephalopathy neuropathological change; HS = hippocampal sclerosis; NACC = National Alzheimer's Coordinating Center; ROSMAP = Religious Orders Study and Rush Memory and Aging Project.

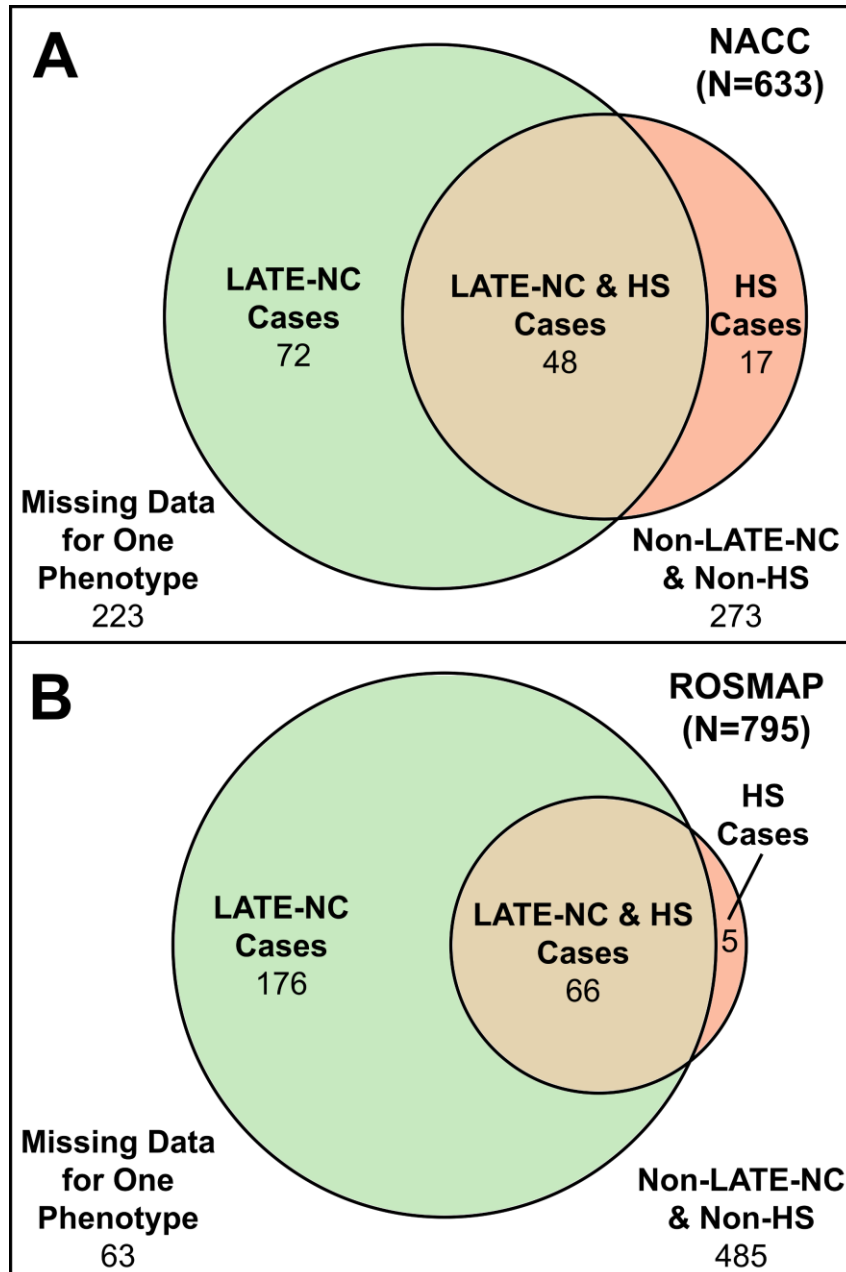


Figure 3.3. Adjusted, meta-analytic, single nucleotide variant (SNV)-level p-values for hippocampal sclerosis (HS) and limbic-predominant age-related TDP-43 encephalopathy neuropathological change (LATE-NC) across **A.** *TMEM106B* ± 10kb and **B.** *GRN* ± 10kb. All analyses were adjusted for sex, age at death, cohort/study, and the first three genetic principal components. Horizontal dashed lines represent the Bonferroni-corrected thresholds for significance that account for the number of independent tests in each genomic region. A diamond represents the SNV with the smallest p-value. The previously identified *TMEM106B* SNV (Rutherford et al., 2012) is labeled and identified with an arrow. MOI = mode of inheritance; LATE-NC = limbic-predominant age-related TDP-43 encephalopathy neuropathological change; HS = hippocampal sclerosis.

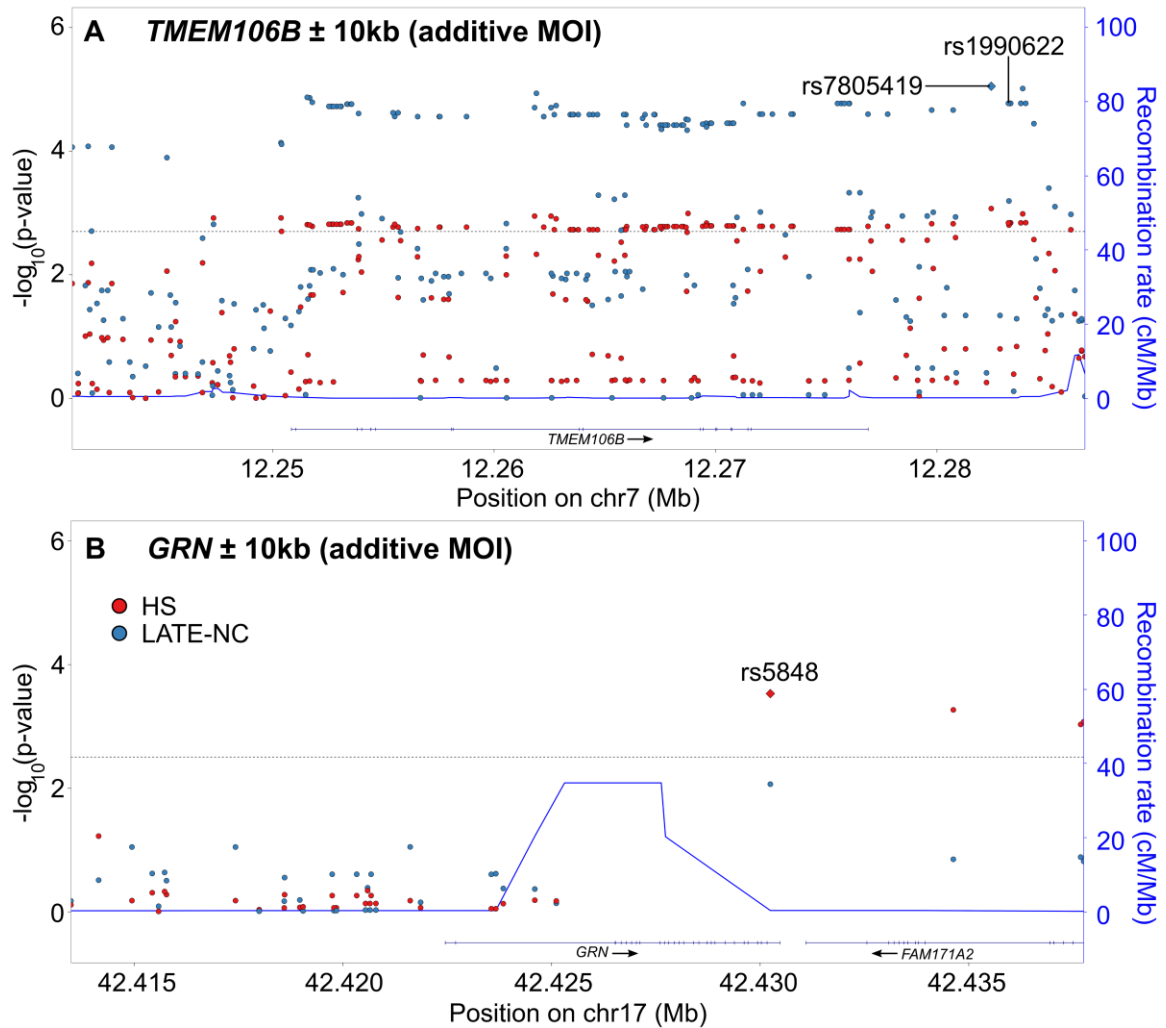


Figure 3.4. Adjusted, meta-analytic, single nucleotide variant (SNV)-level p-values for hippocampal sclerosis (HS) and limbic-predominant age-related TDP-43 encephalopathy neuropathological change (LATE-NC) across *APOE* \pm 10kb. All analyses were adjusted for sex, age at death, cohort/study, and the first three genetic principal components. The horizontal dashed line represents the Bonferroni-corrected threshold for significance that accounts for the number of independent tests in the *APOE* \pm 10kb region. A diamond represents the SNV with the smallest p-value. MOI = mode of inheritance; LATE-NC = limbic-predominant age-related TDP-43 encephalopathy neuropathological change; HS = hippocampal sclerosis.

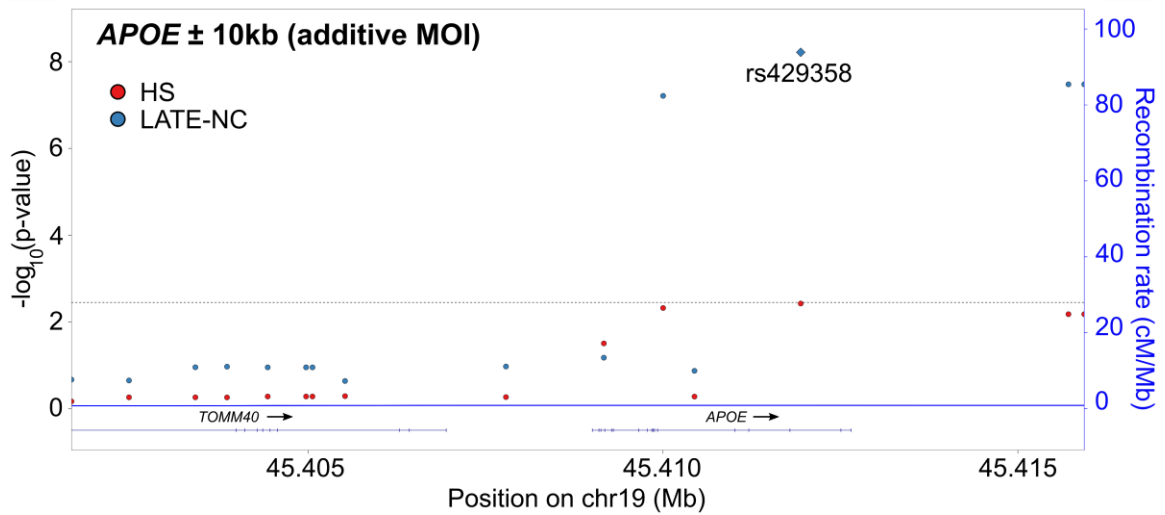


Figure 3.5. Adjusted, meta-analytic, single nucleotide variant (SNV)-level p-values for hippocampal sclerosis (HS) and limbic-predominant age-related TDP-43 encephalopathy neuropathological change (LATE-NC) across *ABCC9* \pm 10kb assuming a recessive mode of inheritance (MOI). A recessive MOI was assumed for *ABCC9* since it has consistently been the MOI with the strongest HS association for *ABCC9* (Nelson et al., 2014; Nelson et al., 2015; Katsumata et al., 2017). All analyses were for sex, age at death, cohort/study, and the first three genetic principal components. The horizontal dashed line represents the Bonferroni-corrected threshold for significance that accounts for the number of independent tests in the *ABCC9* \pm 10kb region. A diamond represents the SNV with the smallest p-value. The previously identified *ABCC9* SNV (Nelson et al., 2014) is labeled and identified with an arrow. MOI = mode of inheritance; LATE-NC = limbic-predominant age-related TDP-43 encephalopathy neuropathological change; HS = hippocampal sclerosis.

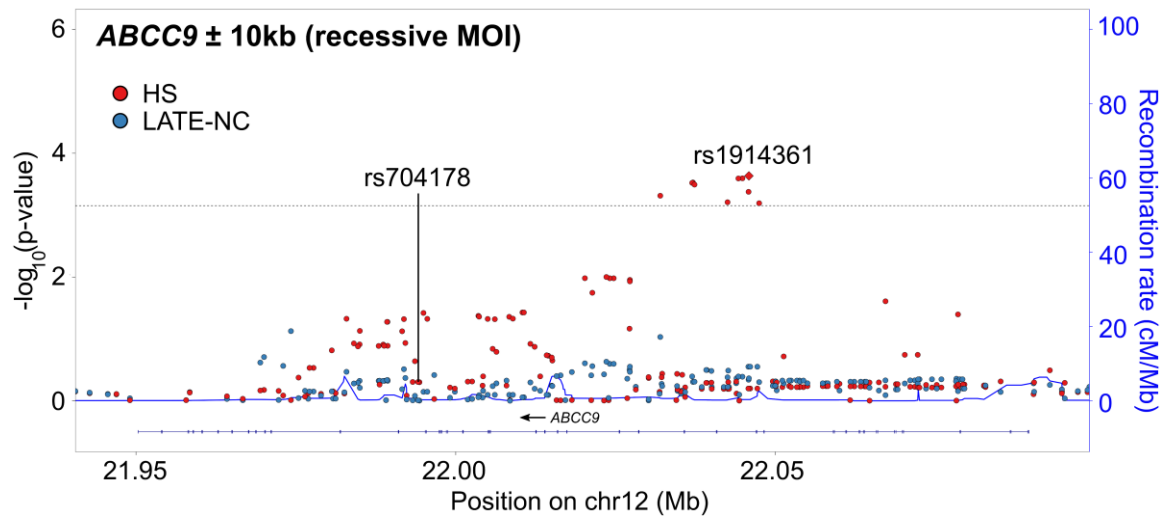


Figure 3.6. Expression quantitative trait loci (eQTL) analyses for rs1914361 and *ABCC9* gene expression across human tissues in the Genotype-Tissue Expression (GTEx) database. **A.** Multi-tissue eQTL plot of rs1914361 and *ABCC9* gene expression; **B.** *ABCC9* normalized gene expression stratified by rs1914361 minor alleles in the nucleus accumbens region of the brain; and **C.** *ABCC9* normalized gene expression stratified by rs1914361 minor alleles in the aorta region of the artery. GTEx = Genotype-Tissue Expression; NES = normalize effect size; eQTL = expression quantitative trait loci.

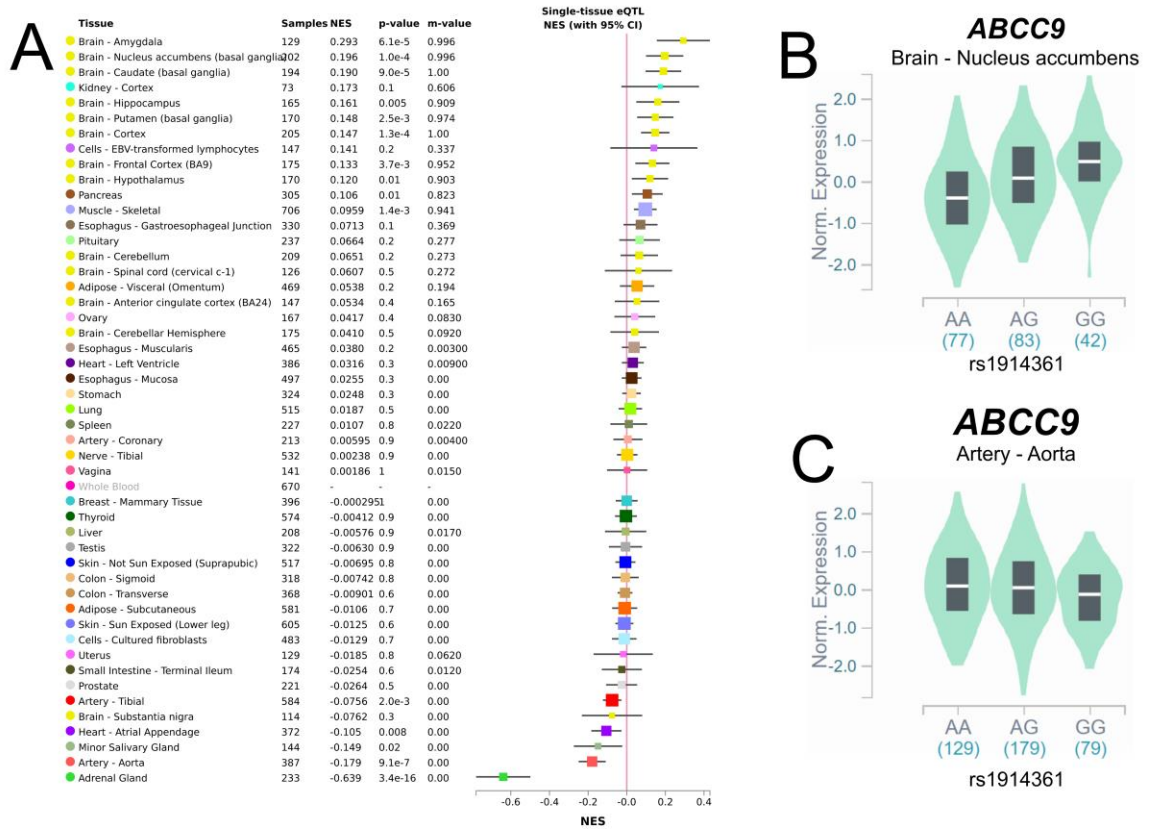


Figure 3.7. Adjusted odds ratio estimates and 95% confidence intervals for genetic single nucleotide variants (SNV) and *APOE* $\epsilon 4$ carrier status from separate regression models of hippocampal sclerosis (HS) fit using data from the National Alzheimer's Coordinating Center (NACC), the Religious Orders Study and Rush Memory and Aging Project (ROSMAP), and the meta-analysis of NACC and ROSMAP. All regression models were adjusted for sex, age at death, cohort/study, and the first three genetic principal components. Some regression models were also adjusted for limbic-predominant age-related TDP-43 encephalopathy neuropathological change (LATE-NC) case status. LATE-NC = limbic-predominant age-related TDP-43 encephalopathy neuropathological change; NACC = National Alzheimer's Coordinating Center; ROSMAP = Religious Orders Study and Rush Memory and Aging Project.

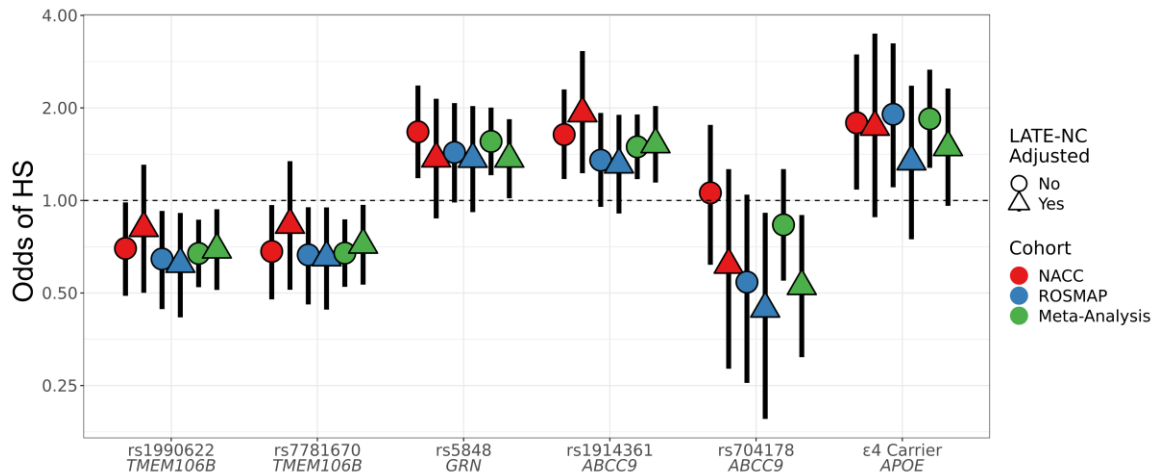


Figure 3.8. *ABCC8* and *ABCC9* gene expression in various human tissues in the **A.** Genotype-Tissue Expression (GTEx) and **B.** Functional Annotation of Human Long Noncoding RNAs via Molecular Mapping (FANTOM5) databases. In GTEx, central nervous system (CNS) tissues included Brodmann (1909) area 24, Brodmann (1909) area 9, C1 segment of cervical spinal cord, amygdala, caudate nucleus, cerebellar hemisphere, cerebellum, cerebral cortex, hippocampus proper, hypothalamus, nucleus accumbens, pituitary gland, and substantia nigra; vascular/smooth muscle tissues included aorta, atrium auricular region, coronary artery, tibial artery, endocervix, esophagus muscularis mucosa, urinary bladder, and uterus; and other tissues included all other tissue types. In FANTOM5, CNS tissues included amygdala, brain, caudate nucleus, cerebellum, diencephalon, dorsal thalamus, globus pallidus, hippocampal formation, locus ceruleus, medulla oblongata, middle frontal gyrus, middle temporal gyrus, occipital cortex, occipital lobe, olfactory apparatus, parietal lobe, pituitary gland, putamen, spinal cord, and substantia nigra; vascular/smooth muscle tissue included artery, heart, heart left ventricle, left cardiac atrium, mitral valve, smooth muscle, tricuspid valve, and uterus; and other tissues included all other tissue types. GTEx = Genotype-Tissue Expression; FANTOM5 = Functional Annotation of Human Long Noncoding RNAs via Molecular Mapping; TPM = transcripts per million; CNS = central nervous system.

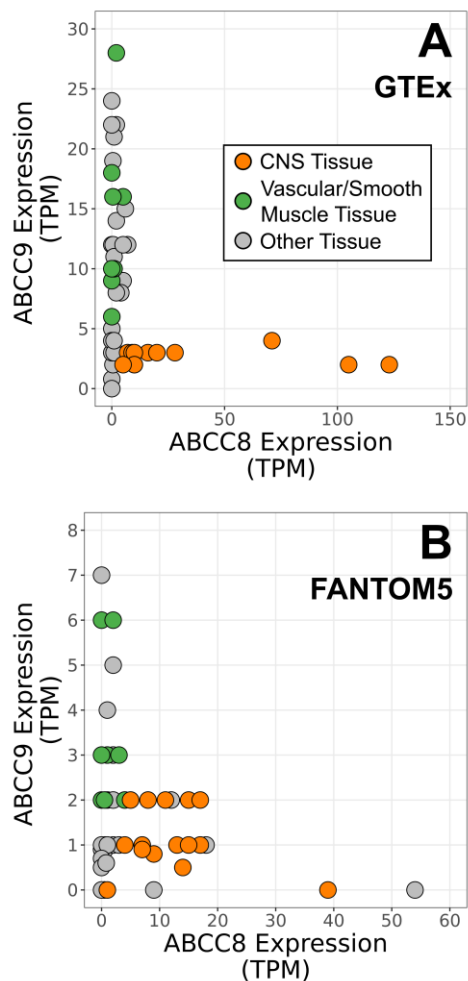
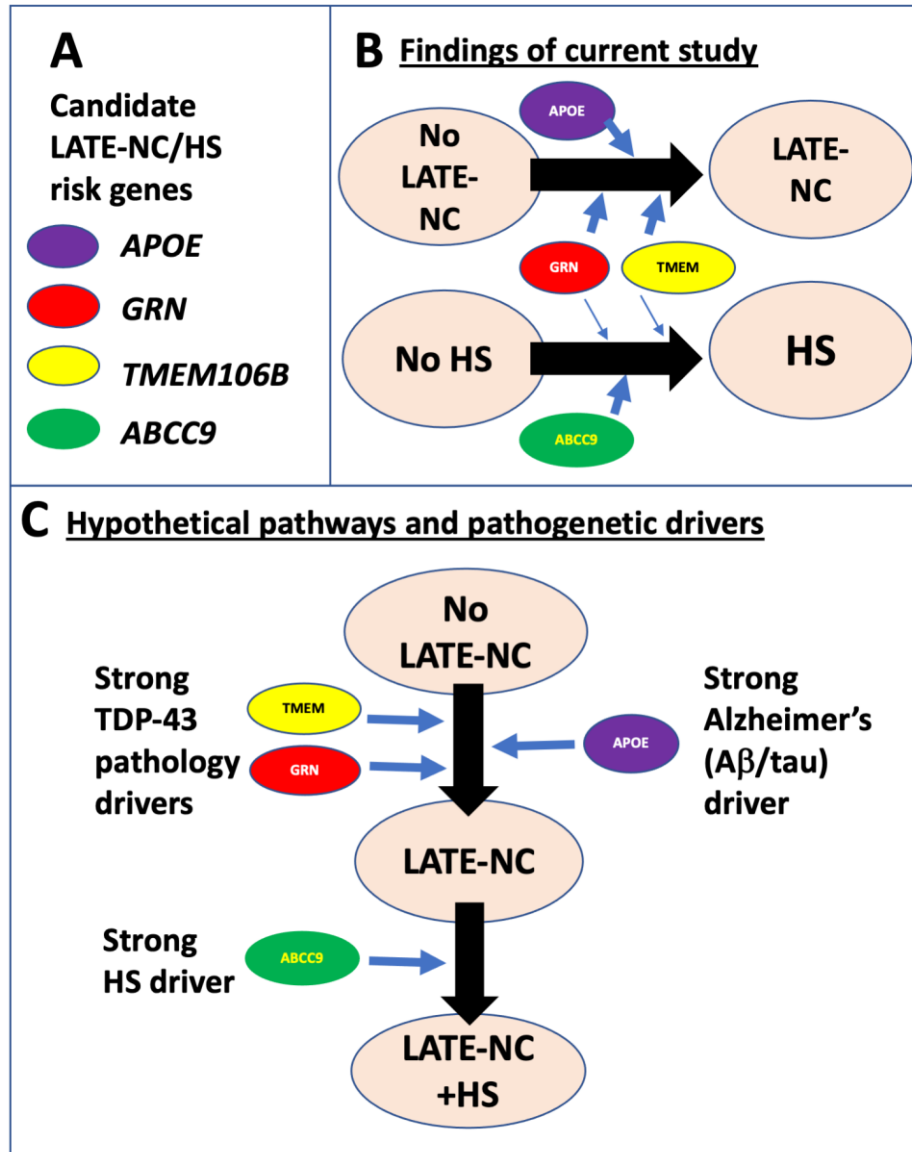


Figure 3.9. Diagrams depicting potential causal relationships between the genes under study with positive findings (*TMEM106B*, *ABCC9*, *GRN*, and *APOE*) and TDP-43 proteinopathy/limbic-predominant age-related TDP-43 encephalopathy (LATE), hippocampal sclerosis (HS), and Alzheimer’s disease (AD). **A.** The candidate genes and their corresponding colors in the diagrams, **B.** a diagram of the current study’s prima facie results, and **C.** a diagram showing hypothetical mechanistic pathways that are compatible with the findings of the current study, including how AD neuropathologic changes (often linked to the *APOE* risk allele) may fit in with the current study’s results. LATE = limbic-predominant age-related TDP-43 encephalopathy; HS = hippocampal sclerosis; AD = Alzheimer’s disease.



CHAPTER 4. A NEW FUNCTIONAL F-STATISTIC FOR GENE-BASED INFERENCE INVOLVING MULTIPLE PHENOTYPES

4.1 Abstract

Genetic pleiotropy is the phenomenon where a single gene or genetic variant influences multiple traits. Numerous statistical methods exist for testing for genetic pleiotropy at the variant level, but fewer methods are available for testing genetic pleiotropy at the gene-level. In the current study, we derive an exact alternative to the Shen and Faraway functional F-statistic for functional-on-scalar regression models. Through extensive simulation studies, we show that this exact alternative performs similarly to the Shen and Faraway F-statistic in gene-based, multi-phenotype analyses and both F-statistics perform better than existing methods in small sample, modest effect size situations. We then apply all methods to real-world, neurodegenerative disease data and identify novel associations.

4.2 Introduction

Genetic pleiotropy is the phenomenon where a single gene or genetic variant influences multiple traits [94]. A recent study estimated that more than half of the human genome contains trait-associated loci and that nearly 90% of those loci are shared by more than one trait [95]. A specific example of pleiotropy in neurodegenerative disease is the *MAPT* gene which is known to contribute risk for several tauopathies and also Parkinson's disease (which is a condition not linked to tau pathology) [49]. Evidence of pleiotropy has also been found across multiple Alzheimer's disease and related dementias (ADRD) traits: AD and Parkinson's disease [45], AD and amyotrophic lateral sclerosis [46], early-onset AD and frontotemporal dementia (FTLD) [47], AD-related psychosis and

schizophrenia [48], and limbic-predominant age-related TDP-43 encephalopathy (LATE) and FTLN-TDP [9].

Simulation studies have found that statistical methods designed to test for associations between a single variant and several phenotypes have higher power than single-phenotype analyses, even when only one of the phenotypes is associated with the variant [96, 97]. Examples of single-variant, multi-phenotype methods include multivariate analysis of variance (MANOVA), TATES [98], mv-BIMBAM [99], SCOPA [100], and MultiPhen [101]. Additionally, gene- and region-based methods have also been shown to have improved power over single-variant approaches [102]. Examples of region-based methods include VEGAS [103], GATES [104], SKAT-O [105], and ACAT [92].

Few statistical methods exist for jointly analyzing multiple phenotypes over genomic regions, though some researchers have tried combining single-variant multiple-phenotype methods with multiple-variant single-phenotype methods with some success [106]. In the current study, we develop a novel test statistic for performing multi-phenotype, gene-based tests by leveraging methods from the branch of statistics known as functional data analysis (FDA). We then compare the performance of the novel test statistic to a similar FDA-based test statistic along with two other methods capable of testing for multi-phenotype, gene-based associations. Finally, we apply the methods to investigate potential pleiotropic effects for several genes and two related but under-studied neurodegenerative diseases, hippocampal sclerosis (HS) and LATE.

4.3 Methods

4.3.1 Functional F-Statistics

4.3.1.1 The Function-on-Scalar Regression Model

In order to utilize methods from functional data analysis (FDA), and the function-on-scalar regression (FoSR) model in particular, a reverse regression approach was used where phenotypes of interest are treated as scalar predictor variables (along with any other adjustors) and genetic information is treated as a functional outcome. Before creating functional outcomes, g minor allele counts were first flipped using the approach of Vsevolzhskaya *et al.* to remove spurious noise [107]. Then the flipped minor allele counts were smoothed for each of the n individuals to create n smooth, individual-level genotype functions, $G_i(t)$, $i \in [1, \dots, n]$. The FoSR model will have the following form:

Equation 1

$$G_i(t) = \beta_0(t) + \beta_1(t)X_{i1} + \dots + \beta_{q-1}(t)X_{i(q-1)} \\ + \beta_q(t)\tilde{Y}_{i1} + \dots + \beta_{p-1}(t)\tilde{Y}_{i(p-q)} + \epsilon_i(t)$$

where $\tilde{Y}_1, \dots, \tilde{Y}_{(p-q)}$ are $(p - q)$ potentially correlated phenotypes, $X_1, \dots, X_{(q-1)}$ are $(q - 1)$ adjustment covariates, and $\beta_j(t)$ is the association function between the genetic region and the j^{th} scalar predictor variable. To test the association between the genotype functions, $G(t)$, and the phenotypes, $\tilde{Y}_1, \dots, \tilde{Y}_{(p-q)}$, we can compare the full model from Equation 1 to the following reduced model:

Equation 2

$$G_i(t) = \beta_0(t) + \beta_1(t)X_{i1} + \dots + \beta_{q-1}(t)X_{i(q-1)} + \epsilon_i(t).$$

4.3.1.2 The Shen and Faraway Functional F-Statistic

A functional F-statistic for FoSR models has been proposed by Shen and Faraway [108] with the following form:

$$F = \frac{(RSS_{Full} - RSS_{Reduced})/(p - q)}{RSS_{Full}/(n - p)}$$

where n is the number of observed genotype functions, p is the total number of parameters in the full model from Equation 1, q is the number of parameters in the reduced model from Equation 2, and RSS are the residual sums of squares of a FoSR model and are defined as follows:

$$RSS = \sum_{i=1}^n \int_{\tau} (G_i(t) - \hat{G}_i(t))^2 dt = \sum_{i=1}^n \sum_{k=1}^{\infty} (G_{ik} - \hat{G}_{ik})^2.$$

Under the null hypothesis, this F-statistic has the following distribution:

Equation 3

$$F \sim \frac{\sum_{k=1}^{\infty} r_k \chi_{p-q}^2 / (p - q)}{\sum_{k=1}^{\infty} r_k \chi_{n-p}^2 / (n - p)}$$

where r_k is the k^{th} decreasingly ordered eigenvalue of the variance-covariance matrix of the genotype functions, $G(t)$. By applying Satterthwaite's approximation, the infinite sums of weighted χ^2 random variables can be approximated by individual χ^2 random variables and the overall distribution can then be approximated by:

$$F \sim \frac{\chi_{f_1}^2 / f_1}{\chi_{f_2}^2 / f_2} = F(f_1, f_2)$$

where $f_1 = c(p - q)$, $f_2 = c(n - p)$, and $c = \frac{\sum_{k=1}^{\infty} r_k^2}{\sum_{k=1}^{\infty} r_k}$.

4.3.1.3 Newly Proposed Functional F-Statistic

Since we expect Satterthwaite's approximation to be imprecise especially for small p-values which are common in genetic association studies, we proposed an alternative derivation. Instead of using Satterthwaite's approximation to simplify the distribution of the functional F-statistic from the ratio of two infinite weighted sums of non-independent χ^2 random variables to the ratio of two independent χ^2 random variables, we propose using known properties of the distributions of quadratic forms to directly evaluate the infinite sums. Specifically, we apply Davies exact method [109, 110] via the `davies()` function from the `CompQuadForm` R package [110] to compute the cumulative distribution function (CDF) values of the infinite sums in the numerator and denominator of Equation 3:

$$(RSS_{Full} - RSS_{Reduced}) \sim \sum_{k=1}^{\infty} r_k^* \chi_{p-q}^2 = CDF_{Numerator}$$

$$RSS_{Full} \sim \sum_{k=1}^{\infty} r_k^* \chi_{n-p}^2 = CDF_{Denominator}$$

where r_k^* is the k^{th} decreasingly ordered eigenvalue of the correlation matrix of the genotype functions, $G(t)$. Correlation matrices were computed from variance-covariance matrices via the `cov2cor()` function in R.

The computed CDF values will be independent and uniformly distributed under the null hypothesis and, thus, can be transformed to follow any distribution. To reflect the traditional F-statistic from multiple linear regression, we transform the numerator and denominator CDF values to χ_{p-q}^2 and χ_{n-p}^2 random variables, respectively. Following this approach, the Shen and Faraway functional F-statistic from Equation 3 will be distributed as $F(p - q, n - p)$ under the null hypothesis.

4.3.1.4 Fitting FoSR Models

Prior to analyzing genomic data with the functional F-statistics, the g minor allele counts were first flipped using the approach of Vsevolozhskaya *et al.* to remove spurious noise [107]. Then, generalized additive models were used to fit penalized cubic regression splines to the flipped minor alleles via the `gam()` function from the `mgcv` R package [111]. See **Figure 4.1** for a comparison of raw and flipped smooth genotype functions, $G_i(t)$, for a region of chromosome 17. Note that the flipped observations no longer necessarily reflect *minor* allele counts, but the underlying associations will be preserved regardless since the choice of effect allele is arbitrary from a statistical standpoint. To reduce the dimensionality of the FoSR models, the individual-level genotype functions, $G_i(t)$, were then evaluated at $g/2$ equally spaced points and these values were subsequently analyzed. This step isn't necessary when dealing with small sample sizes or smaller genetic regions as other simulations we conducted found it to have little effect on statistical power.

FoSR models were fit to the $g/2$ equally spaced $G(t)$ values using linear mixed effects models via the `nlme` R package [112, 113]. The residual sums of squares were then calculated for both the reduced, intercept-only model and the full model with all phenotypes included as predictors. The empirical variance-covariance matrix of the genotype functions, $G(t)$, were calculated using the `var.fd()` function from the `fda` R package [114]. Functional F-statistics were then computed as described earlier.

4.3.2 The Gene Association with Multiple Traits Test

The Gene Association with Multiple Traits (GAMuT) test is a statistical method for cross-phenotype analysis using a nonparametric distance-covariance approach that

compares similarity in multivariate phenotypes to similarity in genotypes across a gene [115]. Briefly, separate similarity matrices are constructed for the phenotypes and genotypes - \mathbf{Y} and \mathbf{X} , respectively - either by projection or through the use of kernel functions. For our simulations and applied analyses, we construct similarity matrices via the projection approach for phenotypes and the linear kernel approach for genotypes. Each similarity matrix is centered to form \mathbf{Y}_c and \mathbf{X}_c and the GAMuT test statistic, T_{GAMuT} , is constructed as follows:

$$T_{GAMuT} = \frac{1}{N} \text{trace}(\mathbf{Y}_c \mathbf{X}_c)$$

where N is the number of individuals included in the analysis. Under the null hypothesis where the two matrices, \mathbf{Y}_c and \mathbf{X}_c , are independent, T_{GAMuT} follows the same asymptotic distribution as

$$\frac{1}{N} \sum_{i,j} \lambda_{X,i} \lambda_{Y,i} z_{ij}^2$$

where $\lambda_{X,i}$ is the i^{th} ordered non-zero eigenvalue of \mathbf{X}_c , $\lambda_{Y,i}$ is the i^{th} ordered non-zero eigenvalue of \mathbf{Y}_c , and z_{ij}^2 are independent and identically distributed χ_1^2 random variables. P-values are then derived using Davies' exact method [109].

All GAMuT analyses were run in R using functions provided by the authors (<https://github.com/epstein-software/GAMuT>) along with their recommended analytic steps (<http://genetics.emory.edu/labs/epstein/software/gamut/GAMuT-example-analysis.html>). Phenotype similarity matrices, \mathbf{Y} , were constructed such that $\mathbf{Y} = \mathbf{P}(\mathbf{P}^T \mathbf{P})^{-1} \mathbf{P}^T$, where \mathbf{P} is a matrix of phenotypes. The raw (i.e., un-flipped and un-smoothed) minor allele counts were used in these analyses. In the applied analyses where

some genotype values were missing for some individuals, variants with any missing values were excluded from the GAMuT analyses.

4.3.3 The Aggregated Cauchy Association Test

The aggregated Cauchy association test (ACAT) is a method that works by converting SNP-level p-values into Cauchy-distributed random variables [92]. Since the sum of dependent Cauchy random variables is identical to the sum of independent Cauchy random variables, no additional information on the linkage disequilibrium or correlation of the SNPs is needed making the method extremely fast [116]. While ACAT was originally developed for the purposes of testing rare variant associations, the authors state that the method can also be applied to common variants.

The ACAT test statistic, T_{ACAT} , has the following form:

$$T_{ACAT} = \sum_{i=1}^k w_i \tan((0.5 - p_i)\pi)$$

where k is the number of SNP-level p-values, p_i is the i^{th} SNP-level p-value, and w_i is the non-negative weight for the i^{th} p-value. Note that under the null hypothesis p_i will be uniformly distributed and $\tan\{(0.5 - p_i)\pi\}$ will be Cauchy distributed [116]. Then, based on the cumulative density function of the Cauchy distribution, the overall ACAT p-value can be approximated by

$$p - \text{value} \approx \frac{1}{2} - \{\arctan(0.5 - p_i)\}$$

where $w = \sum_{i=1}^k w_i$ [92].

In our analyses, we first fit regression models to each SNP and phenotype using linear regression models assuming an additive mode of inheritance via the `lm()` function in R. The raw (i.e., un-flipped and un-smoothed) minor allele counts were used. Then, we applied ACAT with equal weights to combine SNP-level p-values into a single gene-level p-value for each phenotype. The phenotype-level ACAT p-values were then combined via the minimum-p method [117] to obtain a single, multi-phenotype, gene-level p-value. In the context of Equation 1, the smallest gene-level, phenotype-specific p-value would be multiplied by the number of phenotypes analyzed, $(p - q)$, to obtain the final p-value. While this approach is expected to be conservative since it fails to take into account the correlation that may exist among phenotypes, it is computationally efficient and provides an analytic baseline for comparing the performance of other statistical methods. All ACAT analyses were run in R using functions provided by the authors (<https://github.com/yaowuliu/ACAT>).

4.3.4 Simulations

Similar to other simulation studies [118], we utilized realistic linkage disequilibrium patterns by using data from the 1000 Genomes Project [119] for a 100 kb region of chromosome 17 which included 12,735 SNPs spanning from the *FGF11* gene to the *NDEL1* gene. The selection of this region of the genome was arbitrary, but we expect the linkage disequilibrium structure of this region to be representative of, and generalizable to, other regions of the genome.

For each simulation, we randomly selected a window of width m Kb (where $m \in \{10, 25, 50\}$) containing g consecutive SNPs for n individuals sampled with replacement

from the 1,092 available individuals (where $n \in \{100, 250, 500\}$). The resulting matrix was defined to be $\mathbf{G}_{n \times g}$ and standard quality control procedures were then applied such as removing SNPs with minor allele frequencies less than 0.01 and near perfect linkage disequilibrium ($r^2 > 0.99$).

Within each window, we then selected c SNPs (where $c \in \{1, 5, 10\}$) to be causally associated with all five phenotypes. Causal effects, $\beta_{g \times 5}$, were simulated for all SNPs from a normal distribution with $\mu = 0$ and

$$\sigma = \begin{cases} \sigma_{causal} & \text{for all causal SNPs} \\ 0 & \text{for all noncausal SNPs} \end{cases}$$

where $\sigma_{causal} \in \{0.05, 0.10, 0.15, 0.25, 0.50, 1.00\}$. Continuous phenotypes, $\tilde{\mathbf{Y}}_{n \times 5}$, were then generated using the following model:

$$\tilde{\mathbf{Y}}_{n \times 5} = \mathbf{G}_{n \times g} \times \beta_{g \times 5} + \mathbf{E}_{n \times 5}$$

where $\mathbf{E}_{n \times 5}$ is a matrix of errors that follows a multivariate normal distribution with $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Sigma} = \sigma_{error} \mathbf{R}$, $\sigma_{error} = 1$, and \mathbf{R} is a 5×5 matrix of the correlations among the five continuous phenotypes. Multivariate normal observations were generated using the `mvrnorm()` function from the MASS R package [120].

Phenotypes were simulated assuming an underlying equicorrelation structure where all off-diagonal values of \mathbf{R} were equal to ρ , where $\rho \in \{0, 0.5\}$. Disturbance was also added to the underlying correlation matrix, \mathbf{R} , for some simulations so that the methods could be systematically tested in situations with unstructured correlation structures. To create a disturbed correlation matrix, a vector was created consisting of five random values from a continuous Uniform($-b, b$), where $b \in \{0, 5\}$. Then, this vector was multiplied by its transpose to create a 5×5 square matrix, \mathbf{B} , of rank 1. Finally, \mathbf{B} was added to \mathbf{R} and

the resulting matrix was converted to a correlation matrix to create a disturbed underlying correlation matrix, \mathbf{R}^* , for simulating phenotypes.

For Type-I error simulations where $\sigma_{causal} = 0$ for all SNPs, a total of 10,000 simulations were run for each scenario. For all power simulations, a total of 1,000 simulations were run for each scenario.

The performance of the functional F-statistics were compared to two other multi-phenotype, gene-based methods: ACAT and GAMuT. For all of the comparison methods, the raw (i.e., not flipped and not smoothed) genetic data were used. SNP-level analyses for ACAT were performed separately for each phenotype assuming an additive mode of inheritance using linear regression models via the `lm()` function in R.

4.4 Results

4.4.1 Simulations

4.4.1.1 Type I Error

All methods had relatively conservative Type I error rates across the simulation scenarios, though they all tended to become more accurate as the sample size, n , and gene size, m , increased. See **Table 4.1** for Type-I error rates at $\alpha = 0.05$ for each method stratified by gene size, number of observations, correlation among phenotypes, and correlation disturbance.

Under simulation scenarios with no correlation disturbance, ACAT tended to have the best control of Type I error especially for smaller sample sizes. However, as the sample size increased and as disturbance was added to the correlation, the functional F-statistics

tended to perform better. Interestingly, whenever disturbance was added, ACAT's Type I error tended to suffer while GAMuT and the functional F-statistics tended to be either unaffected or become more accurate. Notably, the Type I error rate of the newly derived functional F-statistic was consistently lower than that of the Shen and Faraway functional F-statistic which made it more conservative in all but one scenario (with a gene size of 25,000 bases, sample size of 500, no correlation among phenotypes, and correlation disturbance).

4.4.1.2 Power

No single method had the highest power across all simulation scenarios, but some general trends were apparent. Overall, power increased for all methods as the sample size, gene size, number of causal variants, and causal effect size increased. Additionally, all methods tended to achieve their lowest power in scenarios where the phenotypes were independent ($\rho = 0$ and $b = 0$) while they tended to achieve their highest power in scenarios with high amounts of correlation ($\rho = 0.5$ and $b = 5$). Differences between the methods were observed based on causal effect sizes and correlation structures.

For modest causal effect sizes ($\sigma_{causal} < 0.25$), the functional F-statistics were the most powerful methods regardless of gene size, number of causal variants, and correlation structure (**Figure 4.2**). The performance of the functional F-statistics were similar to one another, though the Shen and Faraway F-statistic tended to have slightly higher power for the smallest causal effect sizes. While ACAT and GAMuT performed worse than the functional F-statistics for modest effect sizes, especially at smaller sample sizes, their performance relative to one another differed substantially based on the underlying

correlation structure. ACAT was far more powerful than GAMuT when the phenotypes were independent ($\rho = 0$ and $b = 0$) and GAMuT was far more powerful than ACAT when there were high amounts of correlation among the phenotypes ($\rho = 0.5$ and $b = 5$). Under their respective most powerful correlation structures, ACAT and GAMuT were able to match the performance of the functional F-statistics at the largest sample size ($N = 500$).

As the the causal effects increased ($\sigma_{causal} \geq 0.25$), ACAT and GAMuT began outperforming the functional F-statistics especially at larger sample sizes (**Figure 4.3**). Similar to the modest effect size scenarios, ACAT tended to be more powerful than GAMuT when the phenotypes were independent ($\rho = 0$ and $b = 0$) and GAMuT tended to be more powerful than ACAT when there were high amounts of correlation among the phenotypes ($\rho = 0$ and $b = 5$). Notably, even with stronger causal effects, the functional F-statistics still tended to be perform similar to, or sometimes even better than, the most powerful method at the smallest sample size. Furthermore, while the relative performances of ACAT and GAMuT varied based on the underlying correlation structure, the functional F-statistics were more stable and tended to be the second and third most powerful methods at larger sample sizes.

4.4.2 Application to Neurodegenerative Disease Data

Several genes have been implicated for hippocampal sclerosis (HS), a neurodegenerative disease characterized by severe neuronal cell loss and gliosis in the hippocampus, including: *KCNMB2*, *TMEM106B*, *ABCC9*, *WWOX*, *GRN*, and *APOE*. Our understanding of HS has evolved in recent years and what was once considered "HS" is

now understood to include several distinct yet related neuropathological diseases, including limbic-predominant age-related TDP-43 encephalopathy (LATE). While genetics are known to play a role in the development of neurodegenerative disorders, the autopsy-based data necessary to definitively diagnose neuropathological changes associated with these conditions is scarce. Using GWAS data from the Alzheimer's Disease Genetics Consortium along with autopsy-derived phenotypes from the National Alzheimer's Coordinating Center (NACC), we tested for joint associations between HS and LATE and the previously identified risk genes [60, 121, 122]. Similar to other studies of NACC participants [58], individuals were excluded if they died prior to age 65, had at least one of 19 rare brain diseases were diagnosed, or if they were missing any adjustment variables or all of the neuropathological endophenotypes being analyzed. Each gene was defined based on the canonical transcripts using the hg19 gene range list from PLINK (<https://www.cog-genomics.org/plink/1.9/resources>) and was flanked by an additional 10kb on both ends to capture potential regulatory variants. ACAT, GAMuT, the Shen-Faraway functional F-statistic, and our newly derived functional F-statistic were applied to the data to test for joint associations between HS and LATE and each gene. Statistical significance was defined to be $p < 0.05$. Note that variants with any missing values were excluded from the GAMuT analyses since it requires the genetic data to have no missingness.

Several genes were found to have a joint association with both HS and LATE, including *TMEM106B*, *GRN*, and *APOE*. The complete results may be found in **Table 4.2**. GAMuT failed to detect any gene-based associations, while both functional F-statistic methods and ACAT were able to detect gene-based associations of HS and LATE with the

TMEM106B and *APOE* genes. Notably, the only method to detect a gene-based association between HS and LATE and *GRN* was the newly derived functional F-statistic.

4.5 Discussion

By leveraging known properties of quadratic forms, we were able to develop an alternative derivation for the distribution of the functional F-statistic for FoSR models. Through simulation studies, we showed that this functional F-statistic performs similarly to the existing Shen and Faraway F-statistic and both F-statistics can outperform other statistical methods designed for testing gene-level pleiotropy. While ACAT and GAMuT tended to be more powerful in situations with stronger causal effects, the functional F-statistics performed better in situations with more modest effect sizes which are more common for complex human diseases.

While the performance of the F-statistics were similar in the current simulation study, it will be important to compare the functional F-statistics in other functional data applications. We expect our newly derived F-statistic to result in more accurate tail probabilities than the Shen and Faraway F-statistic since circumvents a degree of freedom approximation with an exact derivation. This property is important in the context of genetic association studies where small p-values are common. Further simulation studies are needed to determine if their performances are comparable in other contexts.

By applying the newly-derived functional F-statistic to neurodegenerative disease data, we were able to detect a gene-based association between HS and LATE and the *GRN* gene. While *GRN* has previously been shown to be associated with HS and, separately, other AD-related phenotypes, this is the first analysis to find a joint association between

both HS and LATE and the *GRN* gene. Given that sample sizes tend to be relatively small in cohorts with autopsy-derived phenotypes like NACC, the functional F-statistics should be favored for gene-based, multi-phenotype tests in these situations going forward.

The original GAMuT paper compared GAMuT to a multivariate functional regression model and found GAMuT to have superior performance [115]. The approach taken to fitting the functional regression models in that study differs from ours in three important ways [123]. First, they did not use a reverse regression approach, meaning the phenotypes were modeled as a multivariate outcome. Since most clinical phenotypes tend to be correlated and typical multivariate methods assume that outcome vectors are independent of one another, we would expect a multivariate modeling approach to be sub-optimal in the presence of correlated phenotypes. Second, since a multivariate regression approach was used, there is no easy way to include mixtures of continuous and categorical phenotypes in the same analysis. Third, while the minor allele counts were smoothed before modeling, they were not flipped prior to smoothing and so there may have been some residual noise in the genotype functions thus obscuring the genomic signals. Given these deficiencies, we are not surprised that the functional regression approach performed so poorly in the context of correlated phenotypes.

The functional F-statistic approaches have several benefits over the other gene-based, multi-phenotype methods. First, since the genomic data are smoothed prior to analyzing, missing genomic data are implicitly imputed and so more genetic variants can be included in the analyses. While ACAT can partially circumvent this issue by taking a complete case approach to each of the single-variant, single-phenotype analyses, GAMuT requires all missing genomic observations to be imputed prior to analyzing. Second, once

the genomic data have been smoothed, the resulting genotype functions, $G(t)$, can be evaluated at a smaller number of points to effectively reduce the dimensionality to the subsequent analyses (specifically, when the number of evaluation points is less than the number of genomic variants). Simulations showed only a marginal reduction in statistical power when the number of evaluation points was half the number of genomic variants. Third, since a reverse regression approach is used for the functional F-statistics, they can easily test for associations with phenotypes of varying types (numeric, categorical, ordinal, splines, etc.). Fourth, by analyzing a gene-based, multi-phenotype association with a single regression model, no multiplicity corrections are needed for single-gene analyses. Thus, the functional F-statistic methods provide a flexible and scalable framework for conducting gene-based, multi-phenotype analyses.

In the derivation of our new F-statistic, we chose to transform the CDF values to F distributions with $(p - q)$ and $(n - p)$ degrees of freedom so that it would align with the F-statistic from multiple linear regression. That choice of degrees of freedom, along with the choice to transform the CDF values to F distributions, was arbitrary. While some limited simulations (not published) found that $(p - q)$ and $(n - p)$ degrees of freedom performed as well as, if not better than, other alternatives, there likely exists more optimal degree of freedom parameters for this F-statistic. Additionally, there may exist more optimal distributions for transforming the CDF values. Further research is needed.

While running the simulations, we came across two novel properties of ACAT [92]. First, the gene-level ACAT p-value will never exceed the lowest SNP-level p-value. So, unlike a p-value combination test like Fisher's method which can produce a gene-level p-value that's smaller than the minimum SNP-level p-value, ACAT acts more like a

multiplicity correction in that the gene-level p-value is closer to geometric mean of the SNP-level p-values. Second, combining several ACAT p-values via ACAT without weights does not always give the same overall p-value as just combining all of the original p-values via ACAT in one pass. Notably, this only occurs when the first-level ACAT p-values consist of varying numbers of p-values and the issue can be mitigated by modifying the weights of the ACAT analyses.

In conclusion, we derived an alternative to the Shen and Faraway F-statistic for FoSR models. In the context of gene-based, multi-phenotype analyses, our newly derived functional F-statistic performed similarly to the Shen and Faraway F-statistic and both F-statistics outperformed other gene-based, multi-phenotype methods specifically in the small sample, modest effect size scenarios which are common in genetic association studies of autopsy-confirmed complex disease phenotypes like dementia. By applying the newly-derived functional F-statistic to real-world data, we were able to identify a novel association between two Alzheimer disease mimics (HS and LATE) and the *GRN* gene. Since our newly derived functional F-statistic is expected to perform better than the Shen and Faraway functional F-statistic with small p-values, it is a promising method for studies of gene-based genetic pleiotropy.

Table 4.1. Type 1 error at $\alpha = 0.05$ for each method stratified by gene size, number of observations, correlation among phenotypes, and correlation disturbance.

Gene Size	No. of Observations	Correlation Among Phenotypes	Correlation Disturbance	ACAT	GAMuT	SF F	New F
10,000	100	0.0	0	0.052	0.034	0.030	0.022
			5	0.034	0.038	0.028	0.019
		0.5	0	0.048	0.035	0.031	0.021
			5	0.037	0.034	0.032	0.024
	250	0.0	0	0.049	0.047	0.038	0.030
			5	0.040	0.041	0.039	0.031
		0.5	0	0.050	0.042	0.036	0.027
			5	0.037	0.043	0.040	0.030
	500	0.0	0	0.054	0.049	0.044	0.036
			5	0.037	0.046	0.040	0.034
		0.5	0	0.051	0.046	0.040	0.034
			5	0.035	0.050	0.041	0.035
25,000	100	0.0	0	0.051	0.032	0.045	0.035
			5	0.041	0.027	0.045	0.038
		0.5	0	0.047	0.028	0.040	0.035
			5	0.037	0.029	0.045	0.038
	250	0.0	0	0.053	0.041	0.044	0.037
			5	0.041	0.043	0.047	0.040
		0.5	0	0.050	0.039	0.049	0.043
			5	0.037	0.040	0.044	0.039
	500	0.0	0	0.053	0.045	0.048	0.044
			5	0.039	0.045	0.054	0.051
		0.5	0	0.050	0.044	0.048	0.042
			5	0.034	0.045	0.048	0.045
50,000	100	0.0	0	0.054	0.023	0.049	0.042
			5	0.038	0.022	0.047	0.041
		0.5	0	0.048	0.026	0.047	0.039
			5	0.036	0.026	0.050	0.043
	250	0.0	0	0.053	0.037	0.050	0.046
			5	0.037	0.034	0.053	0.048
		0.5	0	0.051	0.035	0.051	0.045
			5	0.037	0.035	0.049	0.045
	500	0.0	0	0.049	0.042	0.050	0.048
			5	0.037	0.038	0.048	0.042
		0.5	0	0.048	0.041	0.050	0.045
			5	0.040	0.040	0.052	0.046

No. = number; ACAT = aggregated Cauchy association test; GAMuT = gene association with multiple traits test; SF F = Shen and Faraway functional F-statistic.

Table 4.2. Gene-based results from the applied analysis looking at the pleiotropic effects of several genes on HS and LATE. P-values < 0.05 are in bold.

Chr.	Gene	Comparison Methods			Functional Analyses						
		No. of SNPs	P-values		No. of Eval. Points	Shen-Faraway F			New F		
			GAMuT	ACAT		F	D.F.	P-value	F	D.F.	P-value
3	KCNMB2	955	0.372	1.000	478	0.370	5.9, 1168.4	0.896	0.115	2.0, 396.0	0.892
7	TMEM106B	255	0.657	0.025	255	3.591	2.6, 510.3	0.018	3.938	2.0, 396.0	0.020
12	ABCC9	293	0.688	0.160	293	1.446	6.3, 1245.5	0.191	2.391	2.0, 396.0	0.093
17	GRN	40	0.903	1.000	40	1.394	4.4, 872.3	0.230	1.221	2.0, 396.0	0.296
19	APOE	38	0.982	0.075	38	1.122	4.4, 868.9	0.346	0.908	2.0, 396.0	0.404

Chr. = chromosome; No. = number; SNP = single nucleotide polymorphism; ACAT = aggregated Cauchy association test; GAMuT = gene association with multiple traits test; Eval. = evaluation; D.F = degrees of freedom.

Figure 4.1. The effect of the minor allele flipping algorithm on the resulting smoothed, individual-level genotype functions, $G_i(t)$.

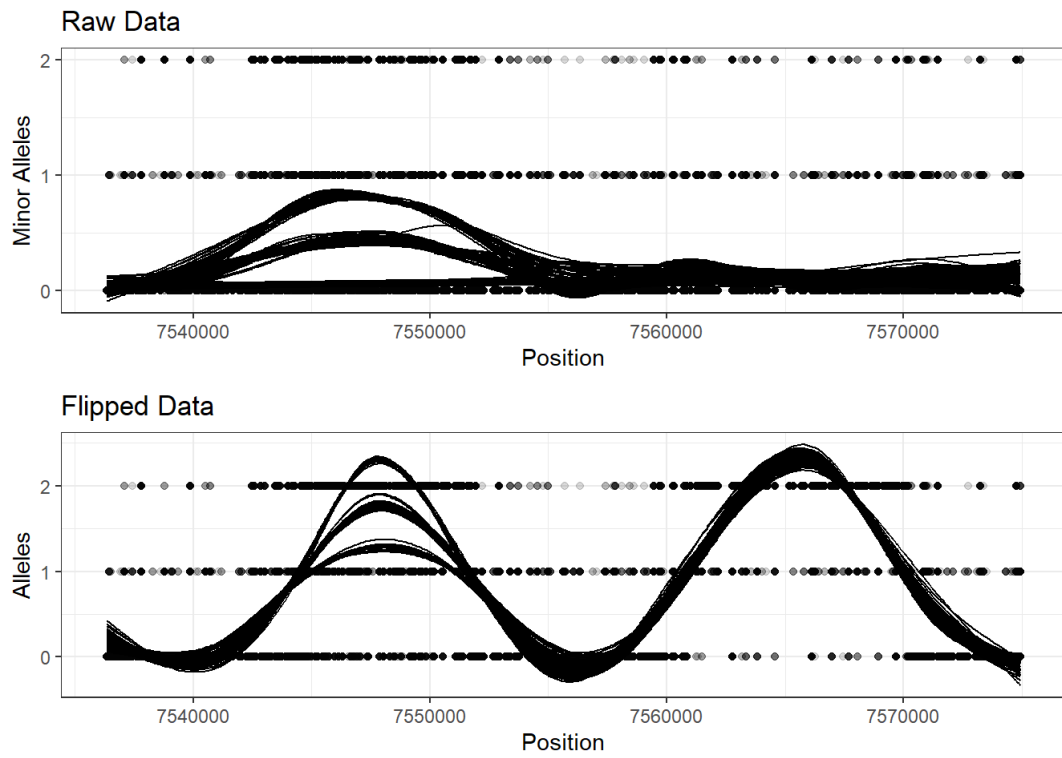
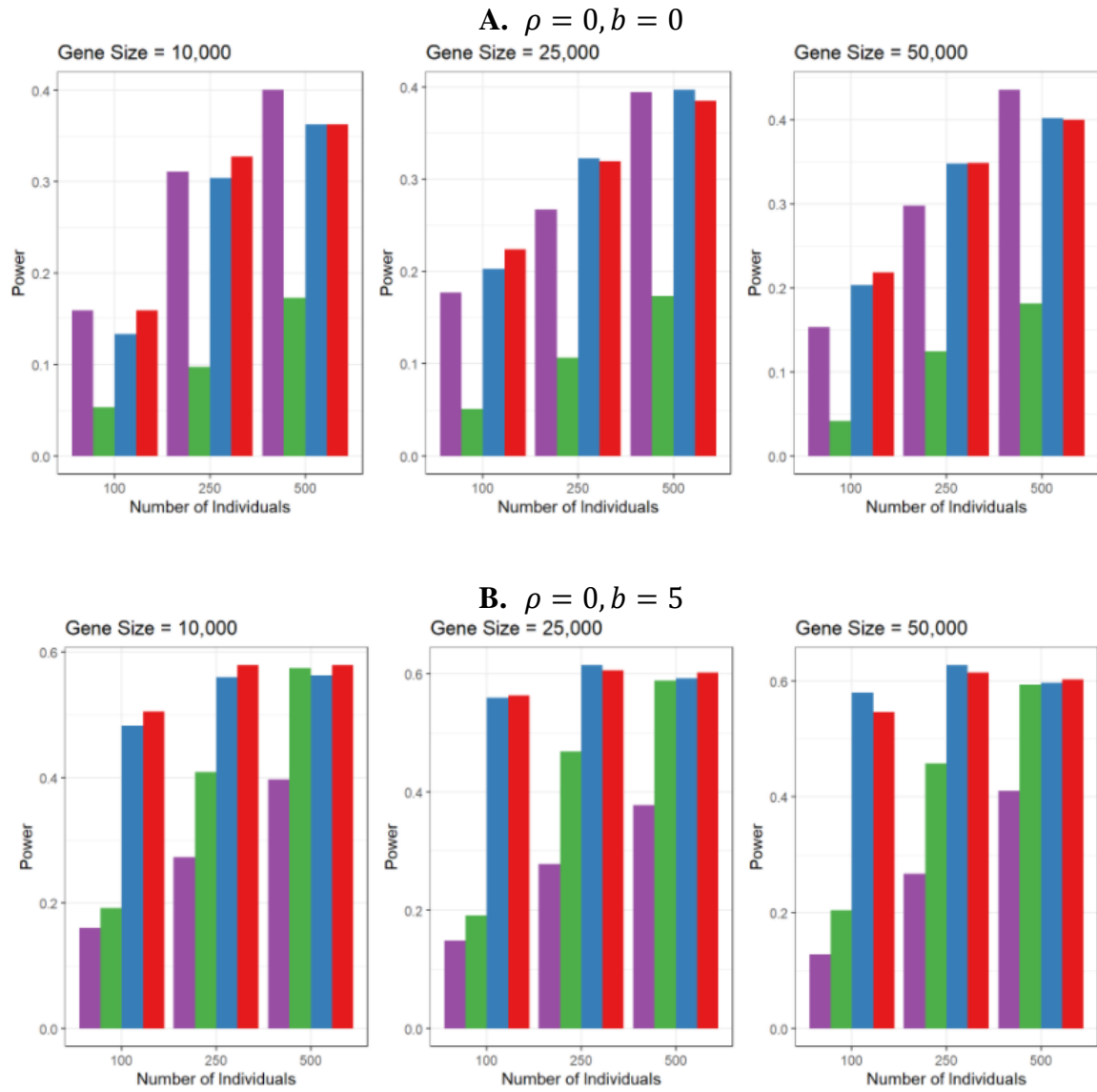
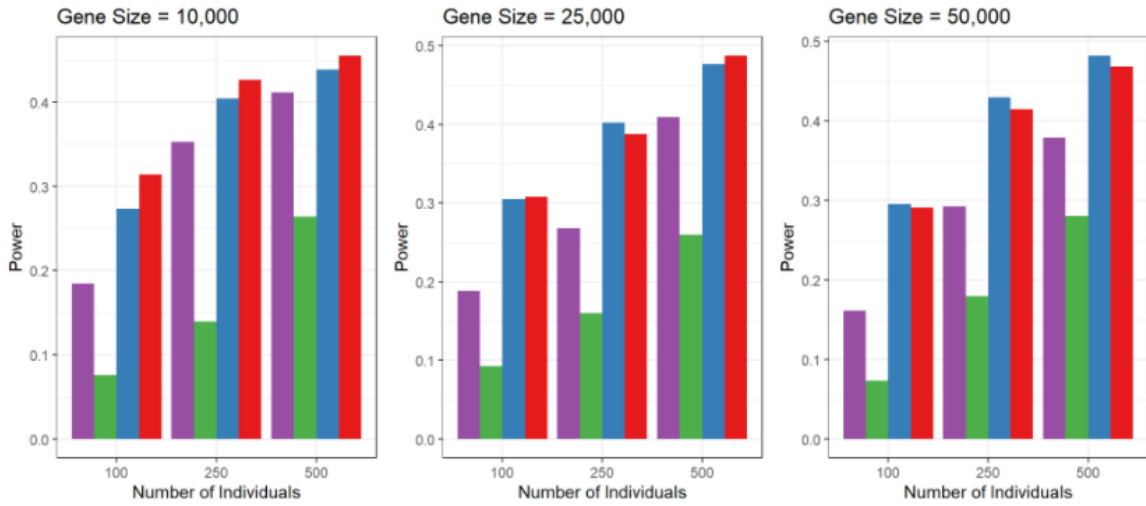


Figure 4.2. Statistical power for simulations assuming 5 continuous phenotypes, 5 causal variants, and causal effect sizes simulated from a normal distribution with $\mu = 0$ and $\sigma_{causal} = 0.15$.



C. $\rho = 0.5, b = 0$



D. $\rho = 0.5, b = 5$

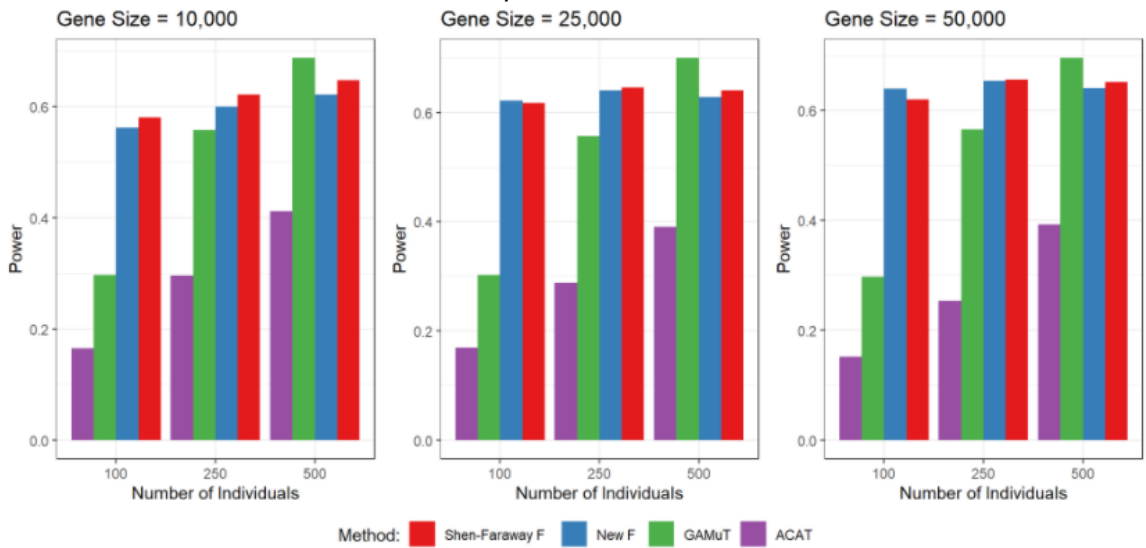
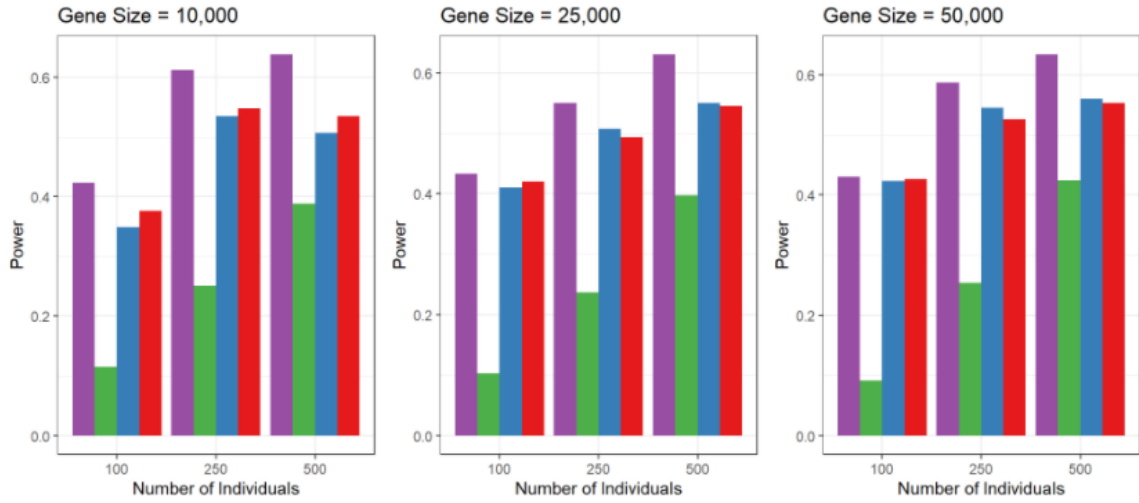
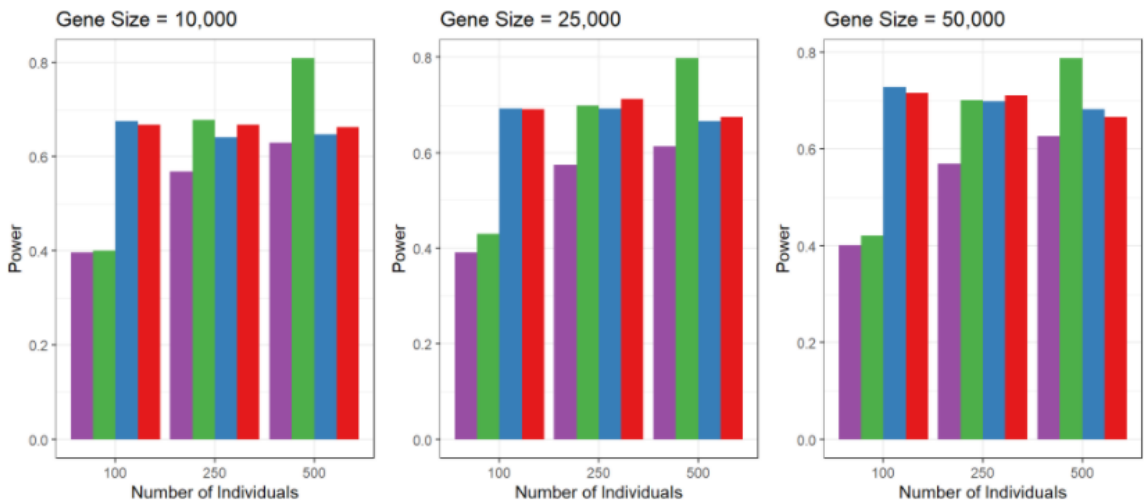


Figure 4.3. Statistical power for simulations assuming 5 continuous phenotypes, 5 causal variants, and causal effect sizes simulated from a normal distribution with $\mu = 0$ and $\sigma_{causal} = 0.25$.

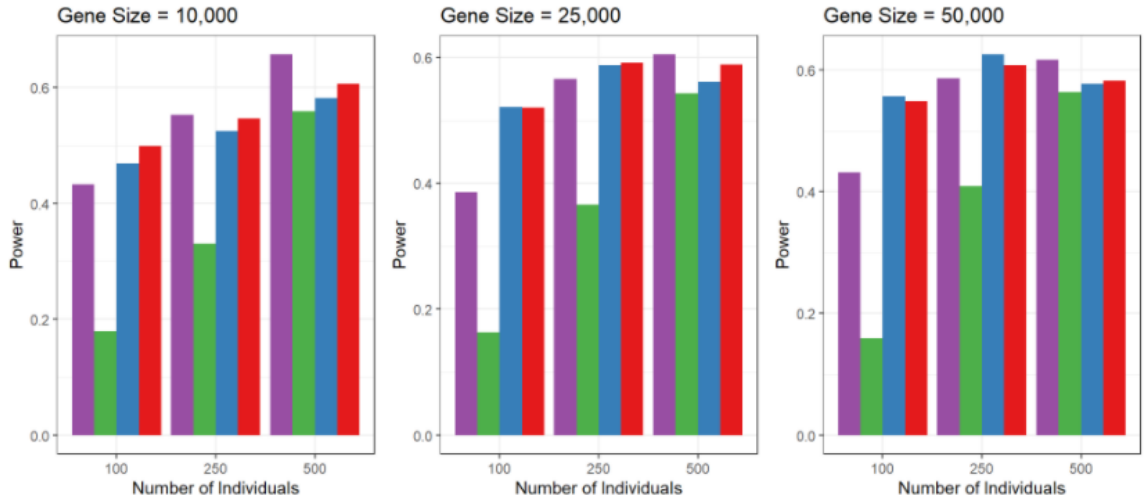
A. $\rho = 0, b = 0$



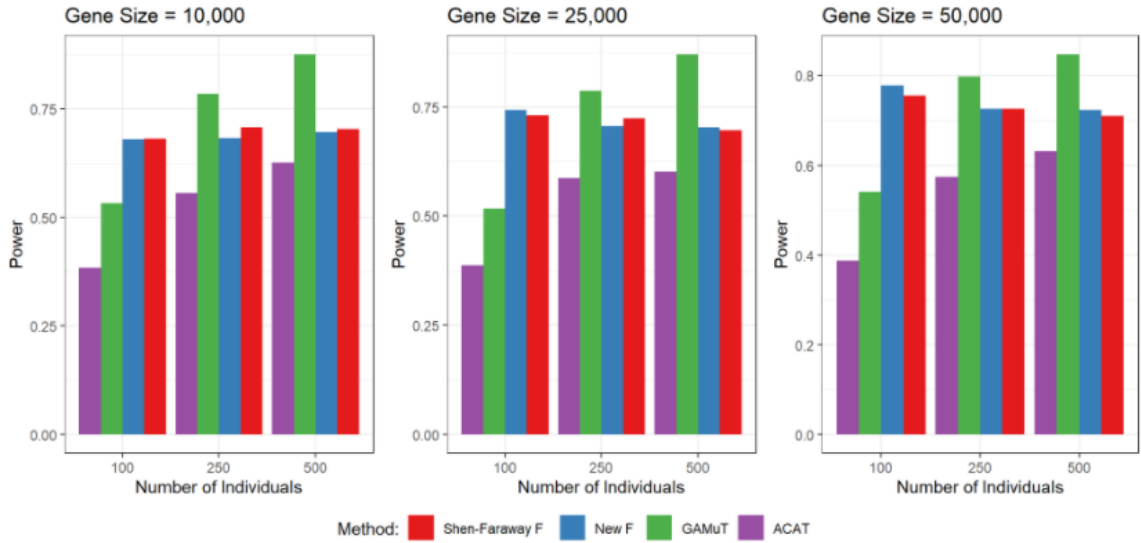
B. $\rho = 0, b = 5$



C. $\rho = 0.5, b = 0$



D. $\rho = 0.5, b = 5$



CHAPTER 5. CONCLUSION

5.1 Summary

Dementia is a complex, multifaceted clinical condition that can be caused by several distinct underlying brain pathologies. While AD is the most common and best-researched cause of dementia, it can only be definitively diagnosed after death via autopsy. Diagnosing AD prior to death via clinical examination has been shown to be inaccurate [124, 125], with one study estimating the sensitivity of clinical diagnoses to range from 70.9% to 87.3% and specificity to range from 44.3% to 70.8% [6]. Despite this, clinically diagnosed AD has been the primary outcome in AD GWASs [18, 19, 21, 22] due to its greater availability in research databases. Since clinical AD diagnoses will almost assuredly contain some false positives and the other AD mimics tend to co-occur, it can be beneficial to investigate clinical AD genetic loci for associations with autopsy-derived neurodegenerative endophenotypes as doing so can help refine our understanding of the genetic associations.

The purpose of this dissertation research was two-fold: 1) to leverage the more specific, autopsy-derived data to better understand the genetics of dementia and 2) develop statistical methodology well suited to these difficult-to-collect, co-occurring phenotypes. The major findings from each of these studies are summarized below.

In the first study, we examined associations between single nucleotide polymorphisms (SNPs) in the *WWOX/MAF* locus and neurofibrillary tangles, neuritic plaques, HS, LATE-NC, and B-ASC. While three recent studies have identified this as an AD risk locus [18, 19, 22], an earlier GWAS identified this same region as an HS risk locus [28] which led us to suspect that this region may be preferentially associated with

non-AD dementia. We found significant associations with HS, LATE-NC, and B-ASC and no significant associations with the AD pathologies (neurofibrillary tangles and neuritic plaques), which suggests that our hunch was not wrong. The associations with LATE-NC and B-ASC have never been reported before and the specific HS SNPs we identified were novel. Additional eQTL analyses found a nominally significant association between the HS SNPs and WWOX expression and nominally significant associations between the LATE-NC and B-ASC SNPs and MAF expression.

The second study used newly collected data to examine associations between several previously identified HS risk genes (*GRN*, *TMEM106B*, *ABCC9*, *KCNMB2*, and *APOE*) and LATE-NC. LATE-NC was found to have significant gene-based associations with *TMEM106B* and *APOE*. HS was also found to have statistically significant gene-based associations with *TMEM106B* and *APOE*, but the LATE-NC associations were stronger suggesting that those genes may favor LATE-NC over HS. The gene-based associations between HS and the *GRN* and *ABCC9* genes were also significant, with the *ABCC9* association being driven by a novel locus within the *ABCC9* gene centered around rs1914361. Importantly, the associations between the *ABCC9* SNPs and HS remained significant and actually produced stronger odds ratio estimates when adjusted for LATE-NC. This observation taken together with the non-significant gene-based association of LATE-NC with *ABCC9* suggests that the association between HS and *ABCC9* is independent of LATE-NC – a surprising finding given how strongly the two pathologies seem to co-occur.

The third study leveraged function-on-scalar regression (FoSR) models from the branch of statistics known as functional data analysis to derive a statistical test for gene-

based, multi-phenotype associations. By treating an individual's allele counts over a genetic region as a "functional" observation and taking a reverse regression approach where the outcome and a predictor are swapped in a regression equation, we were able to apply the FoSR model and setup a gene-based, multi-phenotype test as one would an F-test in traditional linear regression. The existing F-statistic for FoSR models [108] relies on Satterthwaite's approximation to derive the distribution of the statistic under the null hypothesis which we expected to perform suboptimally for small p-values which are common in genetic association studies. With our approach, we instead used Davies' method to directly derive the distribution of the statistic under the null hypothesis. We then compared the two functional F-statistics to competing methods and found that they had superior performance in situations with smaller sample sizes and weaker genetic effects – situations very common for studies investigating genetic associations with autopsy-derived neuropathological data. We then applied the methods to test for joint associations between HS and LATE-NC and the genes investigated in the first two studies (*WWOX*, *GRN*, *TMEM106B*, *ABCC9*, *KCNMB2*, and *APOE*) and identified a significant association with the *TMEM106B* gene.

REFERENCES

1. Arvanitakis, Z., R.C. Shah, and D.A. Bennett, *Diagnosis and Management of Dementia: Review*. JAMA, 2019. **322**(16): p. 1589-1599.
2. Querfurth, H.W. and F.M. LaFerla, *Alzheimer's disease*. N Engl J Med, 2010. **362**(4): p. 329-44.
3. Reitz, C., C. Brayne, and R. Mayeux, *Epidemiology of Alzheimer disease*. Nature reviews. Neurology, 2011. **7**(3): p. 137-152.
4. DeTure, M.A. and D.W. Dickson, *The neuropathological diagnosis of Alzheimer's disease*. Mol Neurodegener, 2019. **14**(1): p. 32.
5. Atri, A., *The Alzheimer's Disease Clinical Spectrum: Diagnosis and Management*. Med Clin North Am, 2019. **103**(2): p. 263-293.
6. Beach, T.G., et al., *Accuracy of the clinical diagnosis of Alzheimer disease at National Institute on Aging Alzheimer Disease Centers, 2005-2010*. Journal of neuropathology and experimental neurology, 2012. **71**(4): p. 266-273.
7. Gaugler, J.E., et al., *Characteristics of patients misdiagnosed with Alzheimer's disease and their medication use: an analysis of the NACC-UDS database*. BMC Geriatr, 2013. **13**: p. 137.
8. Rahimi, J. and G.G. Kovacs, *Prevalence of mixed pathologies in the aging brain*. Alzheimer's research & therapy, 2014. **6**(9): p. 82-82.
9. Nelson, P.T., et al., *Limbic-predominant age-related TDP-43 encephalopathy (LATE): consensus working group report*. Brain, 2019. **142**(6): p. 1503-1527.
10. Karanth, S., et al., *Prevalence and Clinical Phenotype of Quadruple Misfolded Proteins in Older Adults*. JAMA Neurology, 2020.
11. Brenowitz, W.D., et al., *Hippocampal sclerosis of aging is a key Alzheimer's disease mimic: clinical-pathologic correlations and comparisons with both alzheimer's disease and non-tauopathic frontotemporal lobar degeneration*. J Alzheimers Dis, 2014. **39**(3): p. 691-702.
12. Blevins, B.L., et al., *Brain arteriolosclerosis*. Acta Neuropathologica, 2020.
13. Nelson, P.T., et al., *Modeling the association between 43 different clinical and pathological variables and the severity of cognitive impairment in a large autopsy cohort of elderly persons*. Brain Pathol, 2010. **20**(1): p. 66-79.
14. Nag, S., et al., *Hippocampal sclerosis and TDP-43 pathology in aging and Alzheimer disease*. Ann Neurol, 2015. **77**(6): p. 942-52.
15. Robinson, A.C., et al., *Pathological Correlates of Cognitive Impairment in The University of Manchester Longitudinal Study of Cognition in Normal Healthy Old Age*. Journal of Alzheimer's Disease, 2018. **64**: p. 483-496.
16. Nelson, P.T., et al., *Hippocampal sclerosis in advanced age: clinical and pathological features*. Brain, 2011. **134**(Pt 5): p. 1506-18.
17. Amador-Ortiz, C., et al., *TDP-43 immunoreactivity in hippocampal sclerosis and Alzheimer's disease*. Ann Neurol, 2007. **61**(5): p. 435-45.
18. Kunkle, B.W., et al., *Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates A β , tau, immunity and lipid processing*. Nature Genetics, 2019. **51**(3): p. 414-430.

19. Kunkle, B.W., et al., *Novel Alzheimer Disease Risk Loci and Pathways in African American Individuals Using the African Genome Resources Panel: A Meta-analysis*. JAMA Neurol, 2020.
20. Witoelar, A., et al., *Meta-analysis of Alzheimer's disease on 9,751 samples from Norway and IGAP study identifies four risk loci*. Scientific reports, 2018. **8**(1): p. 18088-18088.
21. Lambert, J.-C., et al., *Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease*. Nature Genetics, 2013. **45**(12): p. 1452-1458.
22. Bellenguez, C., et al., *New insights on the genetic etiology of Alzheimer's and related dementia*. medRxiv, 2020: p. 2020.10.01.20200659.
23. Liu, J.Z., Y. Erlich, and J.K. Pickrell, *Case-control association mapping by proxy using family history of disease*. Nat Genet, 2017. **49**(3): p. 325-331.
24. Jansen, I.E., et al., *Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk*. Nature Genetics, 2019. **51**(3): p. 404-413.
25. Dickson, D.W., M. Baker, and R. Rademakers, *Common variant in GRN is a genetic risk factor for hippocampal sclerosis in the elderly*. Neurodegener Dis, 2010. **7**(1-3): p. 170-4.
26. Hokkanen, S.R.K., et al., *Putative risk alleles for LATE-NC with hippocampal sclerosis in population-representative autopsy cohorts*. Brain Pathol, 2020. **30**(2): p. 364-372.
27. Katsumata, Y., et al., *Gene-based association study of genes linked to hippocampal sclerosis of aging neuropathology: GRN, TMEM106B, ABCC9, and KCNMB2*. Neurobiology of aging, 2017. **53**: p. 193.e17-193.e25.
28. Nelson, P.T., et al., *ABCC9 gene polymorphism is associated with hippocampal sclerosis of aging pathology*. Acta neuropathologica, 2014. **127**(6): p. 825-843.
29. Nelson, P.T., et al., *TDP-43 proteinopathy in aging: Associations with risk-associated gene variants and with brain parenchymal thyroid hormone levels*. Neurobiol Dis, 2019. **125**: p. 67-76.
30. Nelson, P.T., et al., *ABCC9/SUR2 in the brain: Implications for hippocampal sclerosis of aging and a potential therapeutic target*. Ageing Res Rev, 2015. **24**(Pt B): p. 111-25.
31. Nelson, P.T., et al., *Reassessment of risk genotypes (GRN, TMEM106B, and ABCC9 variants) associated with hippocampal sclerosis of aging pathology*. Journal of neuropathology and experimental neurology, 2015. **74**(1): p. 75-84.
32. Yang, H.S., et al., *Evaluation of TDP-43 proteinopathy and hippocampal sclerosis in relation to APOE ε4 haplotype status: a community-based cohort study*. Lancet Neurol, 2018. **17**(9): p. 773-781.
33. Rutherford, N.J., et al., *TMEM106B risk variant is implicated in the pathologic presentation of Alzheimer disease*. Neurology, 2012. **79**(7): p. 717-8.
34. Van Langenhove, T., J. van der Zee, and C. Van Broeckhoven, *The molecular basis of the frontotemporal lobar degeneration-amyotrophic lateral sclerosis spectrum*. Ann Med, 2012. **44**(8): p. 817-28.
35. Murray, M.E., et al., *Differential clinicopathologic and genetic features of late-onset amnesic dementias*. Acta Neuropathol, 2014. **128**(3): p. 411-21.

36. Rademakers, R., et al., *Common variation in the miR-659 binding-site of GRN is a major risk factor for TDP43-positive frontotemporal dementia*. Hum Mol Genet, 2008. **17**(23): p. 3631-42.
37. Beecham, G.W., et al., *Genome-wide association meta-analysis of neuropathologic features of Alzheimer's disease and related dementias*. PLoS Genet, 2014. **10**(9): p. e1004606.
38. Robinson, J.L., et al., *Neurodegenerative disease concomitant proteinopathies are prevalent, age-related and APOE4-associated*. Brain, 2018. **141**(7): p. 2181-2193.
39. Chornenky, Y., D.W. Fardo, and P.T. Nelson, *Tau and TDP-43 proteinopathies: kindred pathologic cascades and genetic pleiotropy*. Lab Invest, 2019. **99**(7): p. 993-1007.
40. Josephs, K.A., et al., *Pathological, imaging and genetic characteristics support the existence of distinct TDP-43 types in non-FTLD brains*. Acta Neuropathol, 2019. **137**(2): p. 227-238.
41. Wennberg, A.M., et al., *Association of Apolipoprotein E ϵ 4 With Transactive Response DNA-Binding Protein 43*. JAMA Neurol, 2018. **75**(11): p. 1347-1354.
42. Chou, S.Y., et al., *Genetic susceptibility for ischemic infarction and arteriolosclerosis based on neuropathologic evaluations*. Cerebrovasc Dis, 2013. **36**(3): p. 181-8.
43. Ighodaro, E.T., et al., *Risk factors and global cognitive status related to brain arteriolosclerosis in elderly individuals*. J Cereb Blood Flow Metab, 2017. **37**(1): p. 201-216.
44. Yip, A.G., et al., *APOE, vascular pathology, and the AD brain*. Neurology, 2005. **65**(2): p. 259-65.
45. Ibanez, L., et al., *Pleiotropic Effects of Variants in Dementia Genes in Parkinson Disease*. Frontiers in Neuroscience, 2018. **12**(230).
46. Montibeller, L. and J. de Belleruche, *Amyotrophic lateral sclerosis (ALS) and Alzheimer's disease (AD) are characterised by differential activation of ER stress pathways: focus on UPR target genes*. Cell stress & chaperones, 2018. **23**(5): p. 897-912.
47. Cochran, J.N., et al., *Non-coding and Loss-of-Function Coding Variants in *TET2* are Associated with Multiple Neurodegenerative Diseases*. The American Journal of Human Genetics, 2020. **106**(5): p. 632-645.
48. Creese, B., et al., *Examining the association between genetic liability for schizophrenia and psychotic symptoms in Alzheimer's disease*. Translational Psychiatry, 2019. **9**(1): p. 273.
49. Lin, M.K. and M.J. Farrer, *Genetics and genomics of Parkinson's disease*. Genome Med, 2014. **6**(6): p. 48.
50. Kośła, K., et al., *The WWOX Gene Influences Cellular Pathways in the Neuronal Differentiation of Human Neural Progenitor Cells*. Frontiers in Cellular Neuroscience, 2019. **13**(391).
51. Chang, H.-T., et al., *WW domain-containing oxidoreductase in neuronal injury and neurological diseases*. Oncotarget, 2014. **5**(23): p. 11792-11799.
52. Dourlen, P., et al., *The new genetic landscape of Alzheimer's disease: from amyloid cascade to genetically driven synaptic failure hypothesis?* Acta Neuropathol, 2019. **138**(2): p. 221-236.

53. Sze, C.I., et al., *Down-regulation of WW domain-containing oxidoreductase induces Tau phosphorylation in vitro. A potential role in Alzheimer's disease.* J Biol Chem, 2004. **279**(29): p. 30498-506.
54. Bellenguez, C., et al., *Large meta-analysis of genome-wide association studies expands knowledge of the genetic etiology of Alzheimer's disease and highlights potential translational opportunities.* medRxiv, 2020: p. 2020.10.01.20200659.
55. Bacchelli, E., et al., *An integrated analysis of rare CNV and exome variation in Autism Spectrum Disorder using the Infinium PsychArray.* Sci Rep, 2020. **10**(1): p. 3198.
56. McClay, J.L., et al., *Genome-wide pharmacogenomic study of neurocognition as an indicator of antipsychotic treatment response in schizophrenia.* Neuropsychopharmacology, 2011. **36**(3): p. 616-26.
57. Campbell, P., et al., *Common genetic variants associated with thyroid function may be risk alleles for Hashimoto's disease and Graves' disease.* Clinical Endocrinology, 2016. **84**(2): p. 278-283.
58. Katsumata, Y., et al., *Distinct clinicopathologic clusters of persons with TDP-43 proteinopathy.* Acta Neuropathologica, 2020.
59. Mahoney, E.R., et al., *Brain expression of the vascular endothelial growth factor gene family in cognitive aging and alzheimer's disease.* Molecular Psychiatry, 2019.
60. McCarthy, S., et al., *A reference panel of 64,976 haplotypes for genotype imputation.* Nat Genet, 2016. **48**(10): p. 1279-83.
61. Dumitrescu, L., et al., *Genetic variants and functional pathways associated with resilience to Alzheimer's disease.* Brain, 2020. **143**(8): p. 2561-2575.
62. Marees, A.T., et al., *A tutorial on conducting genome-wide association studies: Quality control and statistical analysis.* International journal of methods in psychiatric research, 2018. **27**(2): p. e1608-e1608.
63. Purcell, S., et al., *PLINK: a tool set for whole-genome association and population-based linkage analyses.* American journal of human genetics, 2007. **81**(3): p. 559-575.
64. R Core Team, *R: A Language and Environment for Statistical Computing.* 2021, R Foundation for Statistical Computing: Vienna, Austria.
65. Balduzzi, S., G. Rücker, and G. Schwarzer, *How to perform a meta-analysis with R: a practical tutorial.* Evidence Based Mental Health, 2019. **22**(4): p. 153-160.
66. Pruim, R.J., et al., *LocusZoom: regional visualization of genome-wide association scan results.* Bioinformatics (Oxford, England), 2010. **26**(18): p. 2336-2337.
67. Machiela, M.J. and S.J. Chanock, *LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants.* Bioinformatics, 2015. **31**(21): p. 3555-7.
68. Gao, X., J. Starmer, and E.R. Martin, *A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms.* Genetic Epidemiology, 2008. **32**(4): p. 361-369.
69. Consortium, G.T., *The Genotype-Tissue Expression (GTEx) project.* Nature genetics, 2013. **45**(6): p. 580-585.
70. Ramasamy, A., et al., *Genetic variability in the regulation of gene expression in ten regions of the human brain.* Nat Neurosci, 2014. **17**(10): p. 1418-1428.

71. Amlie-Wolf, A., et al., *INFERNO: inferring the molecular mechanisms of noncoding genetic variants*. Nucleic Acids Research, 2018. **46**(17): p. 8740-8753.
72. Wang, X., et al., *C-type lectin-like receptor 2 and zonulin are associated with mild cognitive impairment and Alzheimer's disease*. Acta Neurologica Scandinavica, 2020. **141**(3): p. 250-255.
73. Porcellini, E., et al., *Alzheimer's disease gene signature says: beware of brain viral infections*. Immunity & Ageing, 2010. **7**(1): p. 16.
74. Meng, D., et al., *A Role of the Podoplanin-CLEC-2 Axis in Promoting Inflammatory Response After Ischemic Stroke in Mice*. Neurotoxicity Research, 2020.
75. Wang, H., et al., *Genome-wide interaction analysis of pathological hallmarks in Alzheimer's disease*. Neurobiol Aging, 2020. **93**: p. 61-68.
76. Chung, J., et al., *Genome-wide pleiotropy analysis of neuropathological traits related to Alzheimer's disease*. Alzheimer's Research & Therapy, 2018. **10**(1): p. 22.
77. Neltner, J.H., et al., *Arteriolosclerosis that affects multiple brain regions is linked to hippocampal sclerosis of ageing*. Brain, 2014. **137**(Pt 1): p. 255-67.
78. Teng, C.C., et al., *Role of WWOX/WOX1 in Alzheimer's disease pathology and in cell death signaling*. Front Biosci (Schol Ed), 2013. **5**: p. 72-85.
79. Beecham, A.H., et al., *Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis*. Nat Genet, 2013. **45**(11): p. 1353-60.
80. Xia, K., et al., *Genome-wide association analysis identifies common variants influencing infant brain volumes*. Transl Psychiatry, 2017. **7**(8): p. e1188.
81. Ou, S.H., et al., *Cloning and characterization of a novel cellular protein, TDP-43, that binds to human immunodeficiency virus type 1 TAR DNA sequence motifs*. J Virol, 1995. **69**(6): p. 3584-96.
82. Cohen, T.J., V.M. Lee, and J.Q. Trojanowski, *TDP-43 functions and pathogenic mechanisms implicated in TDP-43 proteinopathies*. Trends Mol Med, 2011. **17**(11): p. 659-67.
83. Guo, L. and J. Shorter, *Biology and Pathobiology of TDP-43 and Emergent Therapeutic Strategies*. Cold Spring Harb Perspect Med, 2017. **7**(9).
84. Neumann, M., et al., *Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis*. Science, 2006. **314**(5796): p. 130-3.
85. Mackenzie, I.R. and R. Rademakers, *The role of transactive response DNA-binding protein-43 in amyotrophic lateral sclerosis and frontotemporal dementia*. Curr Opin Neurol, 2008. **21**(6): p. 693-700.
86. Dickson, D.W., et al., *Hippocampal sclerosis: a common pathological feature of dementia in very old (> or = 80 years of age) humans*. Acta Neuropathol, 1994. **88**(3): p. 212-21.
87. Nelson, P.T., et al., *Hippocampal sclerosis of aging, a prevalent and high-morbidity brain disease*. Acta Neuropathologica, 2013. **126**(2): p. 161-177.
88. Ighodaro, E.T., et al., *Hippocampal Sclerosis of Aging Can Be Segmental: Two Cases and Review of the Literature*. J Neuropathol Exp Neurol, 2015. **74**(7): p. 642-52.

89. Aoki, N., et al., *Hippocampal sclerosis in Lewy body disease is a TDP-43 proteinopathy similar to FTLN-TDP Type A*. Acta Neuropathol, 2015. **129**(1): p. 53-64.
90. Wickham, H., *ggplot2: Elegant Graphics for Data Analysis*. 2016: Springer-Verlag New York.
91. Ramey, J.A., *clusteval: Evaluation of Clustering Algorithms*. 2012.
92. Liu, Y., et al., *ACAT: A Fast and Powerful p Value Combination Method for Rare-Variant Analysis in Sequencing Studies*. American journal of human genetics, 2019. **104**(3): p. 410-421.
93. Bryan, J., et al., *ABCC8 and ABCC9: ABC transporters that regulate K⁺ channels*. Pflugers Arch, 2007. **453**(5): p. 703-18.
94. Gratten, J. and P.M. Visscher, *Genetic pleiotropy in complex traits and diseases: implications for genomic medicine*. Genome Medicine, 2016. **8**(1): p. 78.
95. Watanabe, K., et al., *A global overview of pleiotropy and genetic architecture in complex traits*. Nat Genet, 2019. **51**(9): p. 1339-1348.
96. Allison, D.B., et al., *Multiple Phenotype Modeling in Gene-Mapping Studies of Quantitative Traits: Power Advantages*. The American Journal of Human Genetics, 1998. **63**(4): p. 1190-1201.
97. Galesloot, T.E., et al., *A Comparison of Multivariate Genome-Wide Association Methods*. PLOS ONE, 2014. **9**(4): p. e95923.
98. van der Sluis, S., D. Posthuma, and C.V. Dolan, *TATES: Efficient Multivariate Genotype-Phenotype Analysis for Genome-Wide Association Studies*. PLOS Genetics, 2013. **9**(1): p. e1003235.
99. Stephens, M., *A Unified Framework for Association Analysis with Multiple Related Phenotypes*. PLOS ONE, 2013. **8**(7): p. e65245.
100. Mägi, R., et al., *SCOPA and META-SCOPA: software for the analysis and aggregation of genome-wide association studies of multiple correlated phenotypes*. BMC Bioinformatics, 2017. **18**(1): p. 25.
101. O'Reilly, P.F., et al., *MultiPhen: Joint Model of Multiple Phenotypes Can Increase Discovery in GWAS*. PLOS ONE, 2012. **7**(5): p. e34861.
102. Kang, G., B. Jiang, and Y. Cui, *Gene-based Genomewide Association Analysis: A Comparison Study*. Current genomics, 2013. **14**(4): p. 250-255.
103. Liu, J.Z., et al., *A versatile gene-based test for genome-wide association studies*. American journal of human genetics, 2010. **87**(1): p. 139-145.
104. Li, M.-X., et al., *GATES: a rapid and powerful gene-based association test using extended Simes procedure*. American journal of human genetics, 2011. **88**(3): p. 283-293.
105. Lee, S., et al., *Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies*. Am J Hum Genet, 2012. **91**(2): p. 224-37.
106. Chung, J., et al., *Comparison of methods for multivariate gene-based association tests for complex diseases using common variants*. European Journal of Human Genetics, 2019. **27**(5): p. 811-823.
107. Vsevolozhskaya, O.A., et al., *Uncovering Local Trends in Genetic Effects of Multiple Phenotypes via Functional Linear Models*. Genetic Epidemiology, 2016. **40**(3): p. 210-221.

108. Shen, Q. and J. Faraway, *AN F TEST FOR LINEAR MODELS WITH FUNCTIONAL RESPONSES*. *Statistica Sinica*, 2004. **14**(4): p. 1239-1257.
109. Davies, R., *The Distribution of a Linear Combination of χ^2 Random Variables*. *Journal of the Royal Statistical Society Series C*, 1980. **29**(3): p. 323-333.
110. Duchesne, P. and P. Lafaye De Micheaux, *Computing the distribution of quadratic forms: Further comparisons between the Liu-Tang-Zhang approximation and exact methods*. *Computational Statistics & Data Analysis*, 2010. **54**(4): p. 858-862.
111. Wood, S.N., *Generalized Additive Models: An Introduction with R*. 2 ed. 2017: Chapman and Hall/CRC.
112. Jose Pinheiro, D.B., Saikat DebRoy, Deepayan Sarkar, R Core Team, *nlme: Linear and Nonlinear Mixed Effects Models*. 2020.
113. Ruppert, D.a.W., M. P. and Carroll, R. J., *Semiparametric Regression*. Cambridge Series in Statistical and Probabilistic Mathematics. 2003, Cambridge: Cambridge University Press.
114. Hooker, J.O.R.a.S.G.a.G., *fda: Functional Data Analysis*. 2020.
115. Broadaway, K.A., et al., *A Statistical Approach for Testing Cross-Phenotype Effects of Rare Variants*. *Am J Hum Genet*, 2016. **98**(3): p. 525-540.
116. Liu, Y. and J. Xie, *Cauchy Combination Test: A Powerful Test With Analytic p-Value Calculation Under Arbitrary Dependency Structures*. *Journal of the American Statistical Association*, 2020. **115**(529): p. 393-402.
117. Simes, R.J., *An Improved Bonferroni Procedure for Multiple Tests of Significance*. *Biometrika*, 1986. **73**(3): p. 751-754.
118. Vsevolozhskaya, O.A., et al., *DOT: Gene-set analysis by combining decorrelated association statistics*. *PLOS Computational Biology*, 2020. **16**(4): p. e1007819.
119. McVean, G.A., et al., *An integrated map of genetic variation from 1,092 human genomes*. *Nature*, 2012. **491**(7422): p. 56-65.
120. Ripley, W.N.V.a.B.D., *Modern Applied Statistics with S*. 2002, New York: Springer.
121. Besser, L., et al., *Version 3 of the National Alzheimer's Coordinating Center's Uniform Data Set*. *Alzheimer Dis Assoc Disord*, 2018. **32**(4): p. 351-358.
122. Besser, L.M., et al., *The Revised National Alzheimer's Coordinating Center's Neuropathology Form—Available Data and New Analyses*. *Journal of Neuropathology & Experimental Neurology*, 2018. **77**(8): p. 717-726.
123. Wang, Y., et al., *Pleiotropy analysis of quantitative traits at gene level by multivariate functional linear models*. *Genet Epidemiol*, 2015. **39**(4): p. 259-75.
124. Gauthreaux, K., et al., *Concordance of Clinical Alzheimer Diagnosis and Neuropathological Features at Autopsy*. *Journal of neuropathology and experimental neurology*, 2020. **79**(5): p. 465-473.
125. Shim, Y.S., et al., *Clinicopathologic study of Alzheimer's disease: Alzheimer mimics*. *Journal of Alzheimer's disease : JAD*, 2013. **35**(4): p. 799-811.

VITA

ADAM JOSEPH DUGAN

EDUCATION

- M.S. Statistics, University of Kentucky, Lexington, KY, USA May 2015
B.S. Mathematics, University of Georgia, Athens, GA, USA May 2011

PROFESSIONAL EXPERIENCE

- Statistician December 2018 – July 2021
Department of Biostatistics, University of Kentucky, Lexington, KY, USA
- Statistician May 2015 – November 2018
Department of Surgery, University of Kentucky, Lexington, KY, USA
- Graduate Teaching Assistant August 2013 – May 2015
Department of Statistics, University of Kentucky, Lexington, KY, USA
- Primary Instructor June 2014 – July 2014
Department of Statistics, University of Kentucky, Lexington, KY, USA

HONORS AND AWARDS

- Outstanding Doctoral Student Award from the University of Kentucky's College of Public Health. April 2021.
- NSF-Funded Travel Award – Workshop on Emerging Data Science Methods for Complex Biomedical and Cyber Data. March 2020 (delayed due to COVID-19).
- Eastern North American Region (ENAR) of the International Biometric Society Conference Student Travel Award from the University of Kentucky's Biostatistics, Epidemiology, and Research Design (BERD) Core. March 2020 (virtual conference due to COVID-19).
- NSF-Funded Harshbarger Travel Award - SRCOS Summer Research Conference, June 2018
- Best Undergraduate Medical Education Research Abstract at the Council of Residency Directors in EM Academic Assembly 2018. Videoconferencing for Third Year Medical Students' Mid-Clerkship Feedback Sessions. San Antonio, TX, April 2018

PEER-REVIEWED PUBLICATIONS

1. Donigian, A., Hensley, P., Goodwin, J., **Dugan, A. J.**, Hanaoka, M., Bylund, J., & Harris, A. M. (2021). A Quality Improvement Initiative: Patient-Specific Factors Contributing to Extended Pacu Length of Stay and Subsequent Cost. *Physician Leadership Journal*, 8(1), 38–45.
2. Lee JT, Slade E, Uyeda J, Steenburg SD, Chong ST, Tsai R, Raptis D, Linnau KF, Chinapuvvula NR, Dattwyler MP, **Dugan A**, Baghdanian A, Flink C, Baghdanian A, LeBedis CA. American Society of Emergency Radiology Multicenter Blunt Splenic Trauma Study: CT and Clinical Findings. *Radiology*. 2021 Feb 2:202917. doi: 10.1148/radiol.2021202917. Epub ahead of print. PMID: 33529133.
3. Lawrence KA, Vogt D, **Dugan AJ**, Nigam S, Slade E, Smith BN. Mental Health and Psychosocial Functioning in Recently Separated U.S. Women Veterans: Trajectories and Bi-Directional Relationships. *Int J Environ Res Public Health*. 2021 Jan 22;18(3):935. doi: 10.3390/ijerph18030935. PMID: 33498982.
4. Garrett B. Hile, Kaitlin L. Musick, **Adam J. Dugan**, Abby M. Bailey, Gavin T. Howington; Occurrence of Hyperbilirubinemia in Neonates Given a Short-term Course of Ceftriaxone versus Cefotaxime for Sepsis. *The Journal of Pediatric Pharmacology and Therapeutics* 1 January 2021; 26 (1): 99–103. doi: <https://doi.org/10.5863/1551-6776-26.1.99>.
5. Nisiewicz MJ, Kapoor H, Fowler KJ, Furlan A, **Dugan AJ**, Owen JW. Improved survival following transarterial radioembolization of infiltrative-appearance hepatocellular carcinoma. *Abdom Radiol (NY)*. 2021 Jan 1. doi: 10.1007/s00261-020-02870-3. Epub ahead of print. PMID: 33385248.
6. Talal Alnabelsi, Rahul Annabathula, Julie Shelton, Marc Paranzino, Sarah Price Faulkner, Matthew Cook, **Adam J. Dugan**, Sethabhisha Nerusu, Susan S. Smyth, Vedant A. Gupta. Predicting in-hospital mortality after an in-hospital cardiac arrest: A multivariate analysis. *Resuscitation Plus*. Volume 4, 2020, 100039, ISSN 2666-5204, <https://doi.org/10.1016/j.resplu.2020.100039>.
7. Nisiewicz MJ, Roberts JM, Dobbs MR, Ajadi EA, Kitzman P, Wolfe M, Elkins K, **Dugan AJ**, Fraser JF. High Prevalence of Moyamoya Syndrome in Appalachia. *Cerebrovasc Dis*. 2020 Oct 7:1-6. doi: 10.1159/000510750. Epub ahead of print. PMID: 33027801.
8. Sexton T, Chalhoub G, Ye S, Morris W, Annabathula R, **Dugan A**, Smyth S. Autotaxin Activity Predicts 30 Day Mortality in Sepsis Patients and Correlates With Platelet Count and Vascular Dysfunction. *Shock*. 2020 Aug 20. doi: 10.1097/SHK.0000000000001569. Epub ahead of print. PMID: 32826822.
9. Annabathula R, **Dugan A**, Bhalla V, Davis GA, Smyth SS, Gupta VA. Value-based assessment of implementing a Pulmonary Embolism Response Team (PERT). *J Thromb Thrombolysis*. 2020 Jun 15. doi: 10.1007/s11239-020-02188-3. Epub ahead of print. PMID: 32542527.
10. Chang YW, Davenport D, **Dugan A**, Patel JA. Significant morbidity is associated with proximal fecal diversion among high-risk patients who undergo colectomy: A NSQIP analysis [published online ahead of print, 2020 May 18]. *Am J Surg*. 2020;S0002-9610(20)30264-6. doi:10.1016/j.amjsurg.2020.05.007.

11. Gupta M, **Dugan A**, Chacon E, Davenport DL, Shah MB, Marti F, Roth JS, Bernard A, Zwischenberger JB, Gedaly R. Detailed perioperative risk among patients with extreme obesity undergoing nonbariatric general surgery. *Surgery*. 2020 May 14. pii: S0039-6060(20)30159-8. doi: 10.1016/j.surg.2020.03.016. [Epub ahead of print] PubMed PMID: 32418709.
12. Burns JC, DeCoster RC, **Dugan AJ**, Davenport DL, Vasconez HC. Trends in the Surgical Management of Lower Extremity Gustilo Type IIIB/IIIC Injuries. *Plast Reconstr Surg*. 2020;146(1):183-189. doi:10.1097/PRS.00000000000006912.
13. Higgins MM, **Dugan A**, Bell JR, Bylund JR, Harris AM, Bhalodi AA. Identifying a break in the chain: An analysis of where ureteroscopy damage occurs in the hospital cycle. *Urology*. 2020 Apr 8. pii: S0090-4295(20)30337-X. doi: 10.1016/j.urology.2020.02.030. [Epub ahead of print] PubMed PMID: 32277991.
14. Vyas KS, DeCoster RC, Burns JC, Rodgers LT, Shrout MA, Mercer JP, Coquillard C, **Dugan AJ**, Baratta MD, Rinker BD, Vasconez HC. Autologous Fat Grafting Does Not Increase Risk of Oncologic Recurrence in the Reconstructed Breast. *Ann Plast Surg*. 2020 Feb 11. doi: 10.1097/SAP.0000000000002285. [Epub ahead of print] PubMed PMID: 32049757.
15. Wayne NB, Davis GA, Macaulay TE, Beavers CJ, Messerli AW, **Dugan A**, Smyth SS. Reduced dose thrombolysis with ultrasound-facilitated catheter-directed administration for acute pulmonary embolism reduces length of stay. *Journal of Thrombosis and Thrombolysis*. 2019 Dec 3. doi: 10.1007/s11239-019-02009-2. [Epub ahead of print] PubMed PMID: 31797241.
16. O'Brien KJ, Snapp KR, **Dugan AJ**, Westgate PM, Gupta N. Risk factors affecting length of stay in patients with deep neck space infection. *Laryngoscope*. 2019 Nov 25. doi: 10.1002/lary.28367. [Epub ahead of print] PubMed PMID: 31763702
17. DeCoster RC, Stout MA, Burns JC, Shrout MA, Wetzel M, **Dugan AJ**, Rinker BD, Butterfield TA, Webster JM, Vasconez HC. Appalachian Status Is a Negative Predictor of Breast Reconstruction Following Breast Cancer Resection. *Ann Plast Surg*. 2019 Sep 10. doi: 10.1097/SAP.0000000000001965. [Epub ahead of print] PubMed PMID: 31513081.
18. DeCoster RC, Bautista RF Jr, Burns JC, **Dugan AJ**, Edmunds RW, Rinker BD, Webster JM, Vasconez HC. Rural-Urban Differences in Breast Reconstruction Utilization Following Oncologic Resection. *J Rural Health*. 2019 Sep 11. doi: 10.1111/jrh.12396. [Epub ahead of print] PubMed PMID: 31508853.
19. Woolum JA, Bailey AM, **Dugan A**, Agrawal R, Baum RA. Evaluation of infection rates with narrow versus broad-spectrum antibiotic regimens in civilian gunshot open-fracture injury. *Am J Emerg Med*. 2019 Jul 23:158358. doi: 10.1016/j.ajem.2019.158358. PubMed PMID: 31402235.
20. Hulou MM, Garcia CR, Slone SA, **Dugan A**, Lei F, Huang B, Pittman T, Villano JL. Comprehensive Review of Cranial Chordomas Using National Databases in the USA. *Clin Oncol (R Coll Radiol)*. 2019 Jul 11. pii: S0936-6555(19)30254-7. doi: 10.1016/j.clon.2019.06.004. [Epub ahead of print] PubMed PMID: 31303332.
21. Guthrie AM, Baum RA, Carter C, **Dugan A**, Jones L, Tackett T, Bailey AM. Use of Intranasal Ketamine in Pediatric Patients in the Emergency Department. *Pediatr Emerg Care*. 2019 Jul 8. doi: 10.1097/PEC.0000000000001863. [Epub ahead of print] PubMed PMID: 31290798.

22. Chacon E, Eman P, **Dugan A**, Davenport D, Marti F, Ancheta A, Gupta M, Shah M, Gedaly R. Effect of operative duration on infectious complications and mortality following hepatectomy. *HPB (Oxford)*. 2019 Jun 19. pii: S1365-182X(19)30539-8. doi: 10.1016/j.hpb.2019.05.001. [Epub ahead of print] PubMed PMID: 31229489.
23. Harris AM, James A, **Dugan A**, Bylund J. Increased Operative Duration in Minimally-Invasive Partial Nephrectomy is Associated with Significantly Increased Risk of 30-Day Morbidity. *J Endourol*. 2019 May 2. doi: 10.1089/end.2019.0233. [Epub ahead of print] PubMed PMID: 31044616.
24. Peard LM, Goodwin J, Hensley P, **Dugan A**, Bylund J, Harris AM. Examining and Understanding Value: The Impact of Preoperative Characteristics, Intraoperative Variables, and Postoperative Complications on Cost of Robotic-Assisted Laparoscopic Radical Prostatectomy. *J Endourol*. 2019 Apr 24. doi: 10.1089/end.2019.0066. [Epub ahead of print] PubMed PMID: 31017013.
25. Jaclyn M Stoffel, Regan A Baum, **Adam J Dugan**, Abby M Bailey; Variability in training, practice, and prioritization of services among emergency medicine pharmacists, *American Journal of Health-System Pharmacy*, Volume 76, Issue Supplement_1, 8 February 2019, Pages S21–S27.
26. Luis F. Acosta, Roberto Galuppo, Catherine R. García, Evelyn Villacorta, **Adam Dugan**, Ana Lía Castellanos, Roberto Gedaly, James T. Lee. Association Between Sarcopenia and AFP Level in Patients Undergoing Liver Transplantation for Hepatocellular Carcinoma. *Journal of Surgical Research*, Volume 238, 2019, Pages 10-15.
27. Acosta LF, Chacon E, Eman P, **Dugan A**, Davenport D, Gedaly R. Risk of Infectious Complications After Simultaneous Gastrointestinal and Liver Resections for Neuroendocrine Tumor Metastases. *J Surg Res*. 2019 Mar;235:244-249. Epub 2018 Nov 1. PubMed PMID: 30691802.
28. De Stefano F, Garcia CR, Gupta M, Marti F, Turcios L, **Dugan A**, Gedaly R. Outcomes in patients with portal hypertension undergoing gastrointestinal surgery: A propensity score matched analysis from the NSQIP dataset. *Am J Surg*. 2019 Apr;217(4):664-669. doi: 10.1016/j.amjsurg.2018.12.008. Epub 2018 Dec 11. PubMed PMID: 30578032.
29. Zhou Z, Mims T, **Dugan A**, Trott T, Sanderson W, Bronner J. Randomized Evaluation of Videoconference Meetings for Medical Students' Mid-clerkship Feedback Sessions. *West J Emerg Med*. 2018;20(1):163-169.
30. Chacon E, Vilchez V, Eman P, Marti F, Morris-Stiff G, **Dugan A**, Turcios L, Gedaly R. Effect of critical care complications on perioperative mortality and hospital length of stay after hepatectomy: A multicenter analysis of 21,443 patients. *Am J Surg*. 2019 Jul;218(1):151-156. doi: 10.1016/j.amjsurg.2018.11.016. Epub 2018 Nov 30. PubMed PMID: 30528789.
31. Luis F. Acosta, Eduardo Chacon, Pedro Eman, **Adam Dugan**, Daniel Davenport, Roberto Gedaly. Risk of Infectious Complications After Simultaneous Gastrointestinal and Liver Resections for Neuroendocrine Tumor Metastases. *Journal of Surgical Research*, Volume 235, 2019, Pages 244-249.

32. Erpelding SG, Hopkins M, **Dugan A**, Liau JY, Gupta S. Outpatient Surgical Management for Acquired Buried Penis. *Urology*, Volume 123, 247 - 251. PubMed PMID: 30312674.
33. Erpelding SG, **Dugan A**, Isharwal S, Strup, S, James A, Gupta S. Cystectomy for benign disease: readmission, morbidity, and complications. *Can J Urol*. 2018 Oct;25(5):9473-9479.
34. Harris A, Hensley P, Goodwin J, **Dugan A**, Bell J, Bhalodi A, Bylund J. Examining and Understanding Value: The Cost of Preoperative Characteristics, Intraoperative variables, and Postoperative Complications of Minimally-Invasive Partial Nephrectomy. *Urology Practice* 2018; ISSN 2352-0779.
35. Boustany AN, Elmaraghi S, Agochukwu N, Cloyd B, **Dugan AJ**, Rinker B. A breast prosthesis infection update: Two-year incidence, risk factors and management at single institution. *Indian J Plast Surg* 2018;51:7-14.
36. Murphy M, Erpelding S, Chishti A, **Dugan A**, Ziada A, Kiessling S. Influence of BMI in nephrolithiasis in an Appalachian pediatric population: A single-center experience. *Journal of Pediatric Urology*, Volume 14, Issue 4, 330.e1 - 330.e8.
37. Lacy J, Johnson S, **Dugan A**, Gupta S. Urethroplasty Practice Patterns of Genitourinary Reconstructive Surgeons. *Urology Practice*. 2018 May. Volume 5, Issue 3, 223–227.
38. Avila, J. , Smith, B. , Mead, T. , Jurma, D. , Dawson, M. , Mallin, M. and **Dugan, A.** (2018), Does the Addition of M-Mode to B-Mode Ultrasound Increase the Accuracy of Identification of Lung Sliding in Traumatic Pneumothoraces?. *J Ultrasound Med*, 37: 2681-2687.
39. Plymale MA, Davenport DL, **Dugan A**, Zachem A, Roth JS. Ventral hernia repair with poly-4-hydroxybutyrate mesh. *Surg Endosc*. 2018 Apr;32(4):1689-1694. PubMed PMID: 28916979.
40. Ruiz J, **Dugan A**, Davenport DL, Gedaly R. Blood transfusion is a critical determinant of resource utilization and total hospital cost in liver transplantation. *Clin Transplant*. 2018 Feb;32(2). PMID: 29193383.
41. Song K, Chakraborty A, Dawson M, **Dugan A**, Adkins B, Doty C. Does the Podcast Video Playback Speed Affect Comprehension for Novel Curriculum Delivery? A Randomized Trial. *West J Emerg Med*. 2018 Jan;19(1):101-105. PubMed PMID: 29383063.
42. Lacy JM, Madden-Fuentes RJ, **Dugan A**, Peterson AC, Gupta S. Short-term Complication Rates Following Anterior Urethroplasty: An Analysis of National Surgical Quality Improvement Program Data. *Urology*. 2018 Jan;111:197-202. PubMed PMID: 28823639.
43. Lacy, J., **Dugan, A.**, Gupta, S. Characteristics and predictors of complications after urethroplasty: Effect of operative duration, length of stay, and use of tissue transfer. *European Urology Supplements*, 2017; 16(3), E486-E487.
44. Acosta LF, Garcia CR, **Dugan A**, Marti F, Davenport D, Gedaly R. Impact of super obesity on perioperative outcomes after hepatectomy: The weight of the risk. *Surgery*. 2017 Nov;162(5):1026-1031. PubMed PMID: 28866313.
45. Morizio KM, Baum RA, **Dugan A**, Martin JE, Bailey AM. Characterization and Management of Patients with Heroin versus Nonheroin Opioid Overdoses:

- Experience at an Academic Medical Center. *Pharmacotherapy*. 2017 Jul;37(7):781-790. PubMed PMID: 28100012.
46. Kim S, Kramer SP, **Dugan AJ**, Minion DJ, Gurley JC, Davenport DL, Ferraris VA, Saha SP. Cost analysis of iliac stenting performed in the operating room and the catheterization lab: A case-control study. *Int J Surg*. 2016 Dec;36(Pt A):1-7. PubMed PMID: 27746156.
 47. Jackson N, **Dugan A**, Davenport D, Daily M, Shah M, Berger J, Gedaly R. Risk factors for increased resource utilization and critical care complications in patients undergoing hepaticojejunostomy for biliary injuries. *HPB (Oxford)*. 2016 Sep;18(9):712-7. PubMed Central PMCID: PMC5011122.
 48. Wade A, **Dugan A**, Plymale MA, Hoskins J, Zachem A, Roth JS. Hiatal Hernia Cruroplasty with a Running Barbed Suture Compared to Interrupted Suture Repair. *Am Surg*. 2016 Sep;82(9):e271-4. PubMed PMID: 27670546.
 49. Webb AL, **Dugan A**, Burchett W, Barnett K, Patel N, Morehead S, Silverberg M, Doty C, Adkins B, Falvo L. Effect of a Novel Engagement Strategy Using Twitter on Test Performance. *West J Emerg Med*. 2015 Nov;16(6):961-4. Central PMCID: PMC4651604.