2018-04-01

Analysis of Whole Exome Sequence Data in Affected Cousin Pairs from High-Risk Alzheimer's Pedigrees

Lyndsay Ann Staley
Brigham Young University

Analysis of Whole Exome Sequence Data in Affected Cousin Pairs from High-Risk Alzheimer's Pedigrees

Lyndsay Ann Staley

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Science

John Kauwe, Chair
Perry Ridge
Stephen Piccolo

Department of Biology

Brigham Young University

ABSTRACT

Analysis of Whole Exome Sequence Data in Affected Cousin Pairs
from High-Risk Alzheimer's Pedigrees

Lyndsay Ann Staley
Department of Biology, BYU
Master of Science

Genetic factors account for about half of Alzheimer's Disease (AD) risk and only about a quarter of that heritability is accounted for by known variants. Family based approaches to understanding AD genetics may be an effective way to identify additional risk factors. Here we report the results of whole exome sequencing (WES) and analyses done on pairs of AD affected cousins from 19 families from the Utah Population Database (UPDB) with a statistical excess of AD risk.

WES variants passing quality control were additionally filtered by population frequency (minor allele < 0.01) and concordance between cousin pairs, resulting in 564 variants shared by at least one pair of cousins. For each of these variants we conducted in depth annotations using Ingenuity Variant Analysis (IVA), Wellderly Data Allele Frequencies, and literature searches. To further aid in variant prioritization we analyzed each variant for association with Age at Onset of AD, AD Risk, CSF AB42, CSF Tau, CSF PTau and Rate of Disease Decline in data from the Alzheimer's Disease Genetics Consortium (ADGC) and from the Knight Alzheimer's disease research center. Statistical analyses were conducted using PLINK.

Twelve variants (rs201665195, rs28933981, rs148294193, rs147599881, rs61729902, rs140129800, rs191804178, rs200290640, rs199752248, rs45541434, rs141402160 and rs140914494) in eight genes (*ABCA7, TTR*, *PELI3, FCHO1, SNAP91, COX6A2, MUC16, PIDD1, SYT5* and *NOTCH3*) were prioritized using a clear pipeline of IVA filters and the additional analysis information. We propose that these genes and variants are the most interesting for follow-up based on current knowledge.

This family-based approach to finding rare AD variants adds to a growing body of research suggesting a role for *NOTCH3* in late-onset AD. This approach replicated two known AD risk variants and also implicated novel putative risk AD variants and genes. These results suggest that further application of this method of using pairs of cousins may result in additional insights into AD genetics and the ability to find novel rare, causal AD variants.

Keywords: Alzheimer's Disease, genetic analysis, whole exome sequence, variant analysis, late-onset AD, NOTCH3

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Alzheimer's Disease (AD) is a devastating neurodegenerative disease that is currently the sixth leading cause of death in the United States, and the only top ten cause of death to not have a disease-altering treatment to cure, prevent or even slow it down [1]. It affects 10% of people over the age of 65, and as our population gets older and people are able to live longer, it is expected that the rate of AD deaths will continue to increase [2]. Pathologically, AD is characterized by amyloid-beta aggregates and neurofibrillary tangles in the brain. Patients with AD slowly lose their memory and their ability to care for themselves until they are no longer the person their friends and family once knew. Caretakers and medical expenses puts a large financial burden on the families of AD patients, along with the emotional turmoil of watching their loved one decline.

Currently, genetic factors account for about half of AD risk and only 31% of that heritability is accounted for by known variants [3]. There are more than 20 distinct loci that have been associated with AD risk from genome-wide association studies (GWAS) and linkage analyses [4], but the majority of AD genetics is still left unexplained. Familial-based studies are a beneficial and resourceful study design in finding rare variants that increase disease risk, as pedigree information and family history can help narrow the possibilities of risk alleles and provide more power to find rare alleles. We report our study done using AD cousin pairs, from 19 separate families, looking for rare, putative causal AD risk variants. Using first and second cousin pairs makes this study unique and interesting because, assuming the phenotype is derived from a shared variant, this approach allows you to narrow the genetic material down to approximately 12.5% and 3.13%, respectively, and keeps the cost to a minimum by only looking at the exomes of two individuals per family. Finding more loci associated with AD risk will help

us solve the genetic puzzle and get us closer to finding a possible treatment for this terrible disease. Better understanding AD genetics would allow us to discern more of the biological processes that begin years before clinical symptoms appear and would give us a chance to treat early and prevent the damage from becoming irreversible.

METHODS

Data

We collected whole exome sequence data from two individuals, either first or second cousins, in 19 separate families from the Utah Population Database (UPDB) with a statistical excess of AD risk. We retained the variants that were shared between at least one cousin pair, and had a minor allele frequency (MAF) of less than 0.1%, to strictly filter for rare variants. This resulted in a list of 564 variants from 527 genes which we used in the rest of our analyses.

Replication Analyses

We completed additional analyses using other datasets, to look for replication of significance of these 564 variants and to find everything that is known about these variants thus far, in hopes of discovering which could be putative causal AD variants.

When a variant or gene was found in more than 1 cousin pair, we labeled it a "Multiple Hit". There were a total of 19 Multiple Hit genes, containing 36 variants. Ten of these variants were considered "Multiple Hit variants", because the exact variants were seen in multiple cousin pairs, rather than just the gene being seen multiple times. No Multiple Hit genes or variants were seen in more than 2 cousin pairs.

We used the Alzheimer's Disease Genetics Consortium (ADGC) [5] data to do a number of

validation analyses. Briefly, ADGC is imputed SNP array data that includes 28,730 subjects, with 11967 males and 16760 females, and 10486 AD cases and 10168 healthy controls. Of the 564 variants, 291 were sequenced in the ADGC data. We ran an AD Risk logistic regression and an Age at Onset of Disease linear regression using PLINK [6]. We included sex, APOE4, and the first ten EIGENSTRAT principal components (PCs) [4, 7] as covariates in both of these regressions. We used ADGC data to run Gene-Based Testing using SKAT [8] to analyze our 19 Multiple Hits genes for AD Risk at the gene level, rather than just by variant. We also collected MAF information about these variants in ADGC AD cases and cognitively healthy controls.

We tested known AD biomarkers to see if any of the variants were associated with changing the levels of either AB42, Tau or PTau. Of the 564 variants, only 12 variants were sequenced in the CSF data. Briefly, the CSF data we used for these analyses were from the Knight-Alzheimer's Disease Research Center at Washington University School of Medicine (Knight ADRC). This included 3,963 subjects, with 1,895 males and 1,675 females, and 1,479 AD cases and 1,370 healthy controls. All samples were genotyped using the Illumina 610 or the Omniexpress chip. We ran 3 linear regressions, one for each of the phenotype of interest, with age, APOE status, sex and 2 PCs [4, 7] as covariates. We used PLINK to run these 3 analyses.

We were also interested to see how frequent our variants were in the Wellderly Dataset [9]. The Wellderly dataset includes elderly citizens who are cognitively healthy, without medical interventions, at the age of 80 and older. Briefly, Whole Genome Sequencing (WGS) was performed using Complete Genomics platform on 600 Wellderly individuals. We collected MAF information about our variants from this dataset. We also constructed a basic filter to look for the pattern of a risk allele. To pass the filter, the frequency of the variant in the Wellderly dataset needed to be less than or equal to the frequency of the variant in ADGC controls and less

frequent in controls than it was in ADGC AD cases (Wellderly <= ADGC Controls < ADGC AD Cases). If the variant passed this filter, it would seem to be following the pattern of being an AD risk allele.

We ran a Rate of Decline analysis, investigating if any of the variants were associated with faster or slower disease decline, in the Rate of Decline dataset [10]. Briefly, this dataset used longitudinal clinical data from ADNI and Knight ADRC at Washington University to calculate polygenic risk scores (PRS). They used this to compare the genetic architecture of AD Risk and AD disease progression, which they calculated from change in clinical dementia rating (CDR) per year. Lastly, we analyzed our variants using RegulomeDB [11] which is an online database that annotates variants with already known and predicted regulatory elements in the human genome. It gives information on regions of DNAase hypersensitivity, transcription factor binding sites, known eQTLs, DNA methylation, and promoter regions that have been biochemically characterized to regulate transcription. They use data from GEO, ENCODE, and published literature and return scores which represent how much supporting data exists for that variant [11].

Pipeline

In order to sift through all the information clearly and concisely, we created a pipeline with specific cut offs and filters. Figure 1 shows an overview of this pipeline, from data collection to the results. Data were additionally analyzed through the use of Ingenuity® Variant Analysis™ software (https://www.qiagenbioinformatics.com/ products/ingenuity-variant-analysis) from QIAGEN, Inc. We gathered the results from all the additional annotation analyses performed,

and entered them as Custom Annotation columns into our samples on Ingenuity Variant Analysis (IVA).

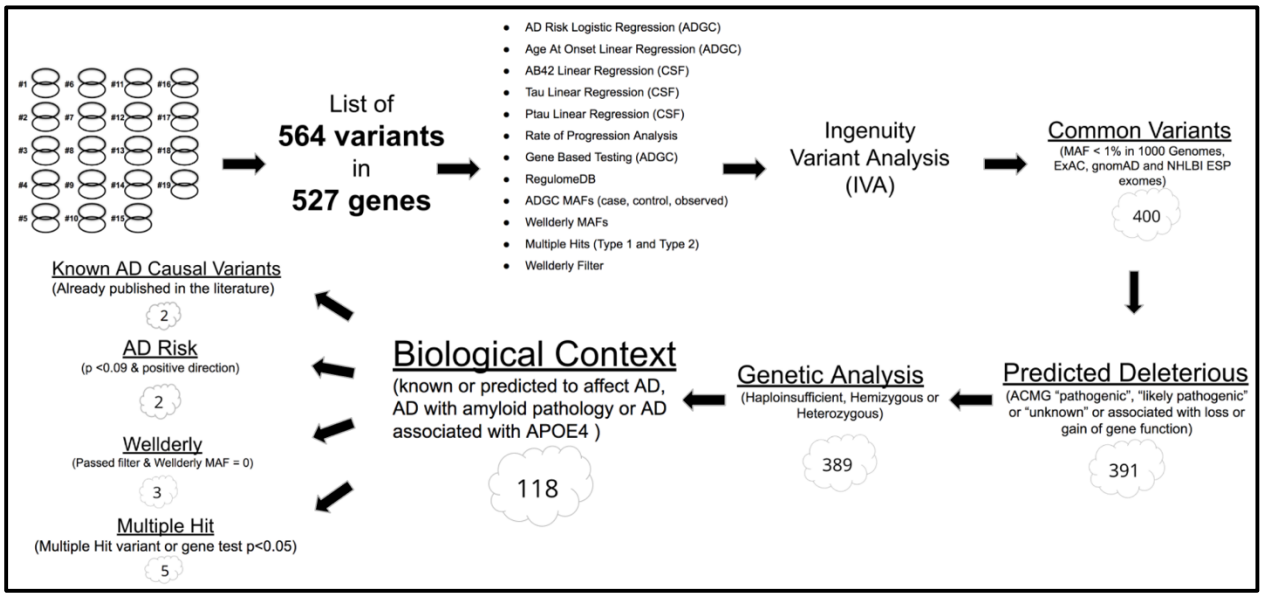

Figure 1. Pipeline Overview: Starting with concordant variants with MAF <0.1% from 19 cousin pairs, (564 variants in 527 genes), we ran a number of validation analyses and completed additional annotation on these variants. We then imputed this information into Ingenuity Variant Analysis (IVA). We filtered using Common Variants, Predicted Deleterious, Genetic Analysis, and Biological Context filters modified to fit our specific study design. Finally, we prioritized variants that passed either our Known AD Causal Variants, AD Risk, Wellderly, or Multiple Hit filters, leaving 12 prioritized variants.

After entering all our information into IVA, we began with the Common Variants filter. We expect a variant found from this type of study to be rare, and not seen frequently in the population. We set the cut off to only include variants with MAF <1% in all the populations IVA offers (1000 Genomes [12], ExAC [13], gnomAD [13], and NHLBI GO Exome Sequencing Project (ESP), Seattle, WA (URL: http://evs.gs.washington.edu/EVS/) [March 2018]). Briefly, 1000 Genomes has 2504 samples from 26 different populations, and their sequencing was generated using Illumina, SOLiD and 454. The Exome Aggregation Consortium (ExAC) seeks to harmonize sequencing data from a variety of large-scale sequencing projects, and 60,706 unrelated individuals sequenced from various populations and disease interest groups are available in this dataset. The Genome Aggregation Database (gnomAD) is an extension of

ExAC, by including 15,496 whole genomes in addition to the 123,136 exome sequence data. Lastly, NHLBI Exome Sequencing Project's (ESP) currently has 6503 samples and their goal is to pioneer the application of next-generation sequencing of the protein coding regions of the human genome across diverse, richly phenotyped populations and to share these datasets and findings with the scientific community. This filter brought the number of variants from 564 down to 400.

The next cut off we used was labeled the Predicted Deleterious filter. We chose for IVA to only keep variants that ACMG [14] considered "Pathogenic", "Likely Pathogenic", or "Unknown", or variants that were associated with loss or gain of gene function. Basically, we wanted to exclude any variants that are already known to not cause a change or not be harmful. This brought the number of variants from 400 to 391. The next step was the Genetic Analysis filter. We chose to only keep variants that were Haploinsufficient, Hemizygous or Heterozygous. This only excluded 2 variants and brought the number of variants down to 389. Then we added the Biological Context filter, which only kept variants known or predicted to affect AD, AD with amyloid pathology, or AD associated with APOE4. This filter brought the variant count from 389 down to 118.

Of these 118 variants that passed all the filters, a variant was added to our final list if it 1) was already published in the literature as a known AD risk variant, 2) had a p value < 0.09 for the AD Risk regression and a positive direction, suggesting it is was following the trend of being a risk allele, 3) passed the Wellderly Filter described previously and had a Wellderly MAF = 0, suggesting that the variant is never seen in healthy elderly population and it could be a risk allele, or 4) was a Multiple Hit variant, where the exact variant was seen in multiple cousin pairs, or had a p-value < 0.05 for the Gene-Based AD Risk test. These 4 final filters, 1) Known AD Risk

Variant, 2) AD Risk, 3) Wellderly and 4) Multiple Hits, narrowed the variant count to 2, 2, 3 and 5 respectively, resulting in 12 final variants.

RESULTS

We believe this pipeline allowed us to best use the information available to us to clearly narrow the list from 564 variants to 12 variants with the strongest cases for AD Risk alleles. Table 1 shows the final list of the 12 prioritized variants.

Table 1. Top 12 Prioritized Variants for AD Risk from the 19 Cousin Pairs

SNP ID	Gene	CHR	Position	Ref	Alt	Final Filter Information
rs201665195	ABCA7	19	1041971	T	G	Known AD Causal Variant
rs28933981	TTR	18	29178610	C	T	Known AD Causal Variant
rs148294193	PELI3	11	66235714	G	C	AD Risk p = 0.05 (+)
rs147599881	FCHO1	19	17881668	G	A	AD Risk p = 0.06 (+)
rs61729902	SNAP91	6	84291977	G	A	Wellderly Filter and MAF = 0
rs140129800	COX6A2	16	31439613	A	C	Wellderly Filter and MAF = 0
rs191804178	MUC16	19	9076546	G	A	Wellderly Filter and MAF = 0
rs200290640	PIDD1	11	800449	G	A	Multiple Hit variant
rs199752248	PIDD1	11	800453	T	C	Multiple Hit variant
rs45541434	SYT5	19	55693469	C	T	Multiple Hit variant
rs141402160	NOTCH3	19	15302615	C	G	Multiple Hit gene, Gene Based Test p = 0.0001
rs140914494	NOTCH3	19	15302857	G	A	Multiple Hit gene, Gene Based Test p = 0.0001

Known AD Risk Variant

The first final filter we used was to check if any of the 118 variants that were already known to be associated with AD in IVA were published as known AD risk variants. Two variants of our variants were already seen in the literature as AD risk variants.

The first variant, rs201665195, is found in the ABCA7 gene on chromosome 19, is a missense mutation, and SIFT [15] labels it "damaging", but it was not sequenced in our validation datasets. Guennec et al 2016 observed this variant, along with other rare ABCA7 variants in AD cases and confirmed that ABCA7 rare, loss of function and predicted damaging missense variants are enriched in patients with AD [16]. Vardarajan et al 2015 also found this rare variant in 2 of their 3 late onset AD (LOAD) cohorts, and this variant was not seen in unaffected individuals [17].

The second variant, rs28933981, is located in the TTR gene on chromosome 18, is a missense mutation, and SIFT labels it "damaging". It was sequenced in our ADGC validation dataset, but it did not pass our significance thresholds. Sassi et al 2016 observed this variant in 6 AD cases, out of their 332 AD cases sample and this variant had a strong effect size (OR = 6.19, 95% CI = 1.099-63.091) [18]. It was also observed in 2 of their 676 controls sample. TTR is a known to be involved in amyloid beta catabolism and has no homologous proteins, suggesting that subtle changes to this protein could be functionally relevant and manifest phenotypically [18].

AD Risk

The conditions for this final filter were 1) the AD Risk logistic regression p-value needed to be less than 0.09, as power was low we were interested in trends of significance for risk and 2) the direction of affect needed to be positive, implying the variant increased the likelihood of having AD. Two variants passed all the previous criteria in the IVA filters, and also met the conditions of the AD Risk final filter.

The third variant, rs148294193, is found in the *PELI3* gene, has a p-value of 0.05 AD Risk with a positive direction, causes a missense mutation, and has been shown in the literature to bind

UBC [19] and APP [20]. UBC has been shown to be significantly decreased in AD brains, regardless of the genotype, suggesting that decreased UBC function may be important in AD pathogenesis, specifically in regard to increased neuronal death and non-regulated APP production [21]. APP is a long-known AD risk gene, and many mutations in this gene are known to cause early-onset AD. APP is cleaved into amyloid beta peptides, which are a major component of the amyloid plaques found in AD patients' brains [22]. PELI3 is a scaffold protein and an intermediate signaling protein in the innate immune response pathway, but no known disorders have been associated with it thus far.

The fourth variant, rs147599881, is located in the *FCHO1* gene, has an AD Risk p-value of 0.06 with a positive direction, causes a missense mutation and also is shown to bind UBC [23] and APP [20]. FCHO1 is involved in vesicle-mediated transportation and clathrin-mediated endocytosis. There are no known disorders associated with this gene or protein yet.

Wellderly

To pass this final filter, variants needed to 1) pass our Wellderly Filter (Wellderly MAF <= ADGC Controls MAF < ADGC AD Cases MAF) and 2) have a MAF = 0 in the Wellderly population. Three variants met these conditions.

The fifth variant, rs61729902, is located in the *SNAP91* gene, causes a missense mutation, and binds UBC [24]. SNAP91 is involved in vesicle-mediated transportation and clathrin-mediated endocytosis. Interestingly, it is a paralog of *PICALM*, which is a known top 10 AD Risk gene [25], and they both have similarities between their functions of clathrin-mediated endocytosis. *SNAP91* is known to be associated with the Cataract 8 disorder.

The sixth variant, rs140129800, is found in the *COX6A2* gene, causes a missense mutation, and binds APP [20]. COX6A2 is a terminal enzyme in the Mitochondrial respiratory chain and there are no known associations with diseases or disorders at this point.

The seventh variant, rs191804178, is located in the *MUC16* gene, causes a missense mutation, and binds UBC [26]. MUC16 is shown to form a protective mucous barrier on the apical surfaces of the epithelia. It is associated with the following disorders: ovarian cancer, endometriosis, psuedo-meigs syndrome, serous cystadenocarcinoma and bronchogenic cyst.

Multiple Hits

To meet the conditions of this final filter, variants must either 1) be a Multiple Hit variant, meaning the same exact variant was found in more than one cousin pair or 2) have a significant Gene-Based AD Risk P-value of less than 0.05. Five variants in three different genes passed these criteria.

The eighth variant and ninth variant, rs200290640 and rs199752248, respectively, are both Multiple Hit variants, lead to splice site loss, and are both found in the *PIDD1* gene, which is shown to bind APOE4 [27]. Rs200290640 is also labeled by SIFT [15] as "damaging". Neither of these variants were sequenced in our validation datasets. *PIDD1* contains a death domain, interacts with other death domain proteins and it is a suggested to be an effector of p53-dependent apoptosis. It is already associated with Poikiloderma with Neutropenia.

The tenth variant, rs45541434, found in the *SYT5* gene, is another Multiple Hit variant, causes a missense mutation, leads to mafb and myf promoter loss, and is shown to bind UBC [24]. This variant was not significant in our validation analyses. SYT5 is thought to act both as negative regulators of vesicle fusion, allowing fusion in the presence of calcium, and as calcium receptors or sensor molecules. No known disorders have been associated with this gene yet.

The eleventh and twelfth variants, rs141402160 and rs140914494, respectively, are both located in the *NOTCH3* gene, cause missense mutations and SIFT labels them "Damaging", but these variants were not sequenced in our validation datasets. NOTCH3 is known to bind PSEN1 [28, 29] and PSEN2 [28]. *NOTCH3* earned a p-value of 0.0001 from the gene-based testing of AD Risk in the ADGC dataset, implying that although these exact variants might not be AD Risk variants, the gene itself is associated with AD Risk. NOTCH3 plays a key role in neural development and is already known to cause the neuronal disease cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL). In 2012, the same region of NOTCH3 we are reporting, was also linked to causing AD in a Turkish family [30].

DISCUSSION

Sassi et al 2018 recently highlighted the difficulties we face in definitive and accurate diagnoses of AD and many Mendelian adult-onset leukodystrophies as many of these progressive neurodegenerative disorders all share similar clinical, neuroimaging and neuropathological features[31]. This study even implied that there is evidence that some of these neurodegenerative disorders, like CADASIL, HDLS and AD, are shades of the same disease spectrum. There have been numerous sources comparing the similarities of AD and CADASIL. Marchesi et al 2014 discussed how both diseases involve damaged small blood vessels, which causes blocked blood flow or impedes the removal of toxic protein aggregates, and how NOTCH3 mutants cause damage to these vessels in the white matter of the brain [32]. This could be the connection between AD and CADASIL's similar pathogenesis [32]. There have also been cases where patients exhibit the coexistence of CADASIL and AD, and it is unclear of whether NOTCH3 mutations are accelerating a pre-existing AD case, or whether AD pathogenesis is accelerating

CADASIL progression [33]. Overall, we believe that NOTCH3 should be pursued as an interesting AD risk gene until the distinction between AD and CADASIL is elucidated.

Overall, we gathered all the information we could find about each variant and gene at this time. Unfortunately, little is known about most variants, genes and proteins, especially the rare variants we are working with here. We are very aware that there could be variants in this list that are putative causal AD Risk variants, that we were not able to identify at this time because of the large amount of variants we are working with, and the relatively little information known about them. Rare variants are seen so infrequently in the population that it is necessary to either have genotype imputation, custom genotyping arrays, or use whole-exome or whole-genome sequencing to access these rare variants [34]. It is difficult to balance the price of deep sequencing and the unsurety of imputation with the number of samples necessary to generate enough power for statistical associations [34]. We fully acknowledge that we may have some false negative variants, based on under-powered analyses. We also acknowledge that because we are only searching the exomes, we could miss interesting hits throughout the genome that are not protein coding. We hope that others are able to use the lists of variants provided, and perform replication analyses to validate whether or not some of these variants are possibly causing AD in these families and other families in the population.

This unique project structure of using first and second cousin pairs has allowed us to pull out rare variants from high risk AD pedigrees in the UPDB. First and second cousins share on average 12.5% and 3.13% of their genetic material, respectively. Assuming the phenotype of interest is derived from a shared variant, the search window is significantly narrowed by using this approach, while still minimizing costs by only sequencing the exomes of two samples per family. We believe that additional projects of this type would be beneficial in finding rare, AD risk

variants in the future and help explain more of the genetic heritability that currently remains a mystery.

FUTURE DIRECTIONS

Of the original 19 cousin pairs from our study, only 8 cousin pairs carry one or more of our 12 prioritized variants. Assuming there is indeed a shared variant or gene that is causing AD in all of these families, these 11 currently unexplained cousin pairs show there is much more to learn from their shared genetic material than we can currently prioritize with our current information. To do more complete variant prioritization, or find variants that are currently considered false negatives, we could go back to the pedigrees, and collect more unaffected and affected samples and look for segregation of variants. If a variant is showing segregation, it would be beneficial to complete functional validation in cell lines to learn more about the biological impact of the interesting variant(s) at the gene expression level, protein function level, and other possible contributions to AD.

It would also be beneficial to analyze these variants in additional AD datasets, like the Alzheimer's Disease Sequencing Project (ADSP) exome data. Briefly, ADSP's fourth release has 15,664 whole exome sequencing data with revised race information available, genotyped using Illumina HiSeq 2000, for case control and family studies [35]. Replicating our Wellderly MAF Filter would be beneficial and interesting to check for patterns of risk alleles in this large dataset and any others AD datasets available. In conclusion, any additional information we can collect about these variants in different populations, datasets, and families, and any biological information we can gather from experimentation would help us make a more complete picture of how these variants are actually affecting their carriers.

CITATIONS

1. Ridge, P.G., M.T. Ebbert, and J.S. Kauwe, *Genetics of Alzheimer's disease.* Biomed Res Int, 2013. **2013**: p. 254954.
2. Association, A.s., *2017 Alzheimer's disease facts and figures.* Alzheimer's & Dementia: The Journal of the Alzheimer's Association, 2017. **13**(4): p. 325-373.
3. Ridge, P.G., et al., *Assessment of the genetic variance of late-onset Alzheimer's disease.* Neurobiol Aging, 2016. **41**: p. 200.e13-20.
4. Lambert, J.C., et al., *Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease.* Nat Genet, 2013. **45**(12): p. 1452-8.
5. Boehme, K., et al., *ADGC 1000 genomes combined workflow (electronic document).* http://kauwelab.byu.edu/Portals/22/adgc_combined_1000G_12032014.pdf, 2014.
6. Chang, C.C., et al., *Second-generation PLINK: rising to the challenge of larger and richer datasets.* Gigascience, 2015. **4**: p. 7.
7. Patterson, N., A.L. Price, and D. Reich, *Population structure and eigenanalysis.* PLoS Genet, 2006. **2**(12): p. e190.
8. Ionita-Laza, I., et al., *Sequence kernel association tests for the combined effect of rare and common variants.* Am J Hum Genet, 2013. **92**(6): p. 841-53.
9. Kauwe, J.S. and A. Goate, *Genes for a 'Wellderly' Life.* Trends Mol Med, 2016. **22**(8): p. 637-9.
10. Del-Aguila, J.L., et al., *Assessment of the Genetic Architecture of Alzheimer's Disease Risk in Rate of Memory Decline.* J Alzheimers Dis, 2018. **62**(2): p. 745-756.
11. Boyle, A.P., et al., *Annotation of functional variation in personal genomes using RegulomeDB.* Genome Res, 2012. **22**(9): p. 1790-7.
12. Auton, A., et al., *A global reference for human genetic variation.* Nature, 2015. **526**(7571): p. 68-74.
13. Lek, M., et al., *Analysis of protein-coding genetic variation in 60,706 humans.* Nature, 2016. **536**(7616): p. 285-91.
14. Richards, S., et al., *Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology.* Genet Med, 2015. **17**(5): p. 405-24.
15. Ng, P.C. and S. Henikoff, *SIFT: Predicting amino acid changes that affect protein function.* Nucleic Acids Res, 2003. **31**(13): p. 3812-4.
16. Le Guennec, K., et al., *ABCA7 rare variants and Alzheimer disease risk.* Neurology, 2016. **86**(23): p. 2134-7.
17. Vardarajan, B.N., et al., *Rare coding mutations identified by sequencing of Alzheimer disease genome-wide association studies loci.* Ann Neurol, 2015. **78**(3): p. 487-98.
18. Sassi, C., et al., *Influence of Coding Variability in APP-Abeta Metabolism Genes in Sporadic Alzheimer's Disease.* PLoS One, 2016. **11**(6): p. e0150079.
19. Giegerich, A.K., et al., *Autophagy-dependent PELI3 degradation inhibits proinflammatory IL1B expression.* Autophagy, 2014. **10**(11): p. 1937-52.
20. Olah, J., et al., *Interactions of pathological hallmark proteins: tubulin polymerization promoting protein/p25, beta-amyloid, and alpha-synuclein.* J Biol Chem, 2011. **286**(39): p. 34088-100.

21. Stieren, E.S., et al., *Ubiquilin-1 is a molecular chaperone for the amyloid precursor protein.* J Biol Chem, 2011. **286**(41): p. 35689-98.
22. Hunter, S. and C. Brayne, *Understanding the roles of mutations in the amyloid precursor protein in Alzheimer disease.* Mol Psychiatry, 2018. **23**(1): p. 81-93.
23. Zhang, X., et al., *An Interaction Landscape of Ubiquitin Signaling.* Mol Cell, 2017. **65**(5): p. 941-955.e8.
24. Na, C.H., et al., *Synaptic protein ubiquitination in rat brain revealed by antibody-based ubiquitome analysis.* J Proteome Res, 2012. **11**(9): p. 4722-32.
25. Rosenberg, R.N., et al., *Genomics of Alzheimer Disease: A Review.* JAMA Neurol, 2016. **73**(7): p. 867-74.
26. Nathan, J.A., et al., *Why do cellular proteins linked to K63-polyubiquitin chains not associate with proteasomes?* Embo j, 2013. **32**(4): p. 552-65.
27. Theendakara, V., et al., *Direct Transcriptional Effects of Apolipoprotein E.* J Neurosci, 2016. **36**(3): p. 685-700.
28. Saxena, M.T., et al., *Murine notch homologs (N1-4) undergo presenilin-dependent proteolysis.* J Biol Chem, 2001. **276**(43): p. 40268-73.
29. Mizutani, T., et al., *Conservation of the biochemical mechanisms of signal transduction among mammalian Notch family members.* Proc Natl Acad Sci U S A, 2001. **98**(16): p. 9026-31.
30. Guerreiro, R.J., et al., *Exome sequencing reveals an unexpected genetic cause of disease: NOTCH3 mutation in a Turkish family with Alzheimer's disease.* Neurobiol Aging, 2012. **33**(5): p. 1008.e17-23.
31. Sassi, C., et al., *Mendelian adult-onset leukodystrophy genes in Alzheimer's disease: critical influence of CSF1R and NOTCH3.* Neurobiol Aging, 2018.
32. Marchesi, V.T., *Alzheimer's disease and CADASIL are heritable, adult-onset dementias that both involve damaged small blood vessels.* Cell Mol Life Sci, 2014. **71**(6): p. 949-55.
33. Thijs, V., et al., *Coexistence of CADASIL and Alzheimer's disease.* J Neurol Neurosurg Psychiatry, 2003. **74**(6): p. 790-2.
34. Bomba, L., K. Walter, and N. Soranzo, *The impact of rare and low-frequency genetic variants in common disease.* Genome Biol, 2017. **18**(1): p. 77.
35. Butkiewicz, M., et al., *Functional Annotation of genomic variants in studies of Late-Onset Alzheimer's Disease.* Bioinformatics, 2018.