Title: Repository of Alzheimer's Disease and Related Dementia (ADRD) Rare Variants and Application to Diverse Populations

Authors: R. Vaza[1], A. Elmas[1], H. Wu[1], V. Subramaniam[1], A. Goate[1,2], P. O'Reilly[1,2], K-l. Huang[1,2]

Summary

Alzheimer's disease (AD) and related dementias (ADRD) present significant public health challenges, and understanding their genetic underpinnings is crucial for developing effective interventions and risk prediction strategies. We introduce the Repository of Alzheimer's Disease and Related Dementia Variants (RADR), a curated database of 3530 rare genetic variants associated with AD and related dementia (ADRD). The database integrates data from various sources, including Alzforum, ClinVar, the Dominantly Inherited Alzheimer Network (DIAN) and DNA-seq studies. We present the challenges of fragmentation in existing genetic resources. Additionally, we explore the distribution and pathogenicity of genetic variants within RADR with a focus on pathogenic, likely pathogenic and risk factor/modifiers. We demonstrate the application of RADR to two large-scale biobanks, BioMe (n = 30813) and the UK Biobank (n = 201,136), for carrier and allele frequency analysis of pathogenic, likely pathogenic, and risk factor/modifier variants. The findings highlight incomplete penetrance, enrichment of certain variants, and potential population-specific differences. Overall, RADR provides a comprehensive and unified resource for researchers and clinicians studying AD and related dementias, facilitating genetic variant analysis and potential to aid in the development of personalized approaches for diagnosis, prevention, and treatment.

Introduction

Alzheimer's disease (AD) represents a significant public health challenge due to its increasing prevalence in society. As of 2023, the prevalence of the disease is approximated at 6.7 million within the United States alone.[1] AD is a highly heritable neurodegenerative disorder that is characterized by the aggregation and deposition of amyloid-B peptides and hyperphosphorylated tau into neurofibrillary tangles, followed by neuronal loss which results in cognitive and functional decline, leading to severe impairment in daily life activities.[2] Understanding the genetic underpinnings of AD holds strong potential for identifying novel therapeutic targets and improving risk prediction.

This paper seeks to contribute to assist in mapping the genetic architecture knowledge by developing a repository that focuses on both rare and common variants associated with Alzheimer's disease and related dementia. Decades of research have established the genetic basis of Alzheimer's disease, yet this complex landscape remains far from fully mapped. Known common genetic risk variants, such as the Apolipoprotein E (APOE) ε4 allele, have a substantial effect on AD risk but cannot account for all genetic predisposition to the disease.[3,4,5] Meanwhile,

rare variants—mutations with minor allele frequencies (MAF) typically less than 1% in a population—have the potential to substantially contribute to disease risk, but their role in AD remains largely unexplored due to the challenges in detecting and interpreting their effects.

Despite the advancement in genome-wide association studies (GWAS), AD genetics research has largely been centered on common variants–mutations with MAF greater than 5% in a population–with less attention given to the rare variants due to their low frequency and the challenges associated with studying them. AD GWAS studies combined have resulted in the identification of 101 independent variants across 81 loci that are associated with AD/dementia, however, most of these variants are common. These common variants account for a limited percentage of the heritability of AD.[6,7] Heritability estimates the fraction of phenotypic variation due to genetic effects. Twin studies estimate that the heritability of AD is between 60-80%. However, summary statistics used by Wightman et al. (2022) indicate that the SNP-heritability of AD is 3.1%. The gap in the heritability between twin studies and those by GWAS studies is referred to as “missing heritability” and can be attributed to several factors. A potential source of this missing heritability is rare variants, which is actively being discovered by ongoing DNA-Sequencing studies including the Alzheimer’s Disease Sequencing Project (ADSP) and the Alzheimer Disease European Sequencing consortium (ADES). For example, a recent collaborative study of ADSP and ADES included 16k AD cases and 16k controls, identifying *ATP8B4*, ABCA1, and ADAM10 (suggestive) in addition to known rare variants in *TREM2, MAPT, ABI3, PLCG2, SORL1, ABCA7, AKAP9, PLXNA4*, and *UNC5C*[13,74–82].

Compared to hundreds of common variants with small effect sizes revealed by GWAS, pathogenic variants—which are often rare and high-penetrance—could directly serve as clinical biomarkers. These variants are hypothesized to be under intense selective pressure and thus remain at low frequency in a population. Early familial linkage studies and recent large DNA-sequencing (DNA-Seq) datasets are quickly expanding the numbers of ADRD-associated variants.[8-17] However, they are scattered across multiple studies and databases, impeding research and clinical use. To develop clinical genetic tests for Alzheimer’s Disease and Related Dementia’s (ADRD) and borrowing experience of other diseases with genetic tests (ex. hereditary cancer, cystic fibrosis), a unified repository is required.

To address the fragmentation of current ADRD genetic resources, we developed RADR, a systematically curated database of common and rare variants (prototyped). RADR integrates data from DIAN, Alzforum, ClinVar, DNA-Seq studies, and annotations from our genomic pipeline, adding pathogenicity classification distinct from commonly used functional scores (ex. CADD, SIFT).

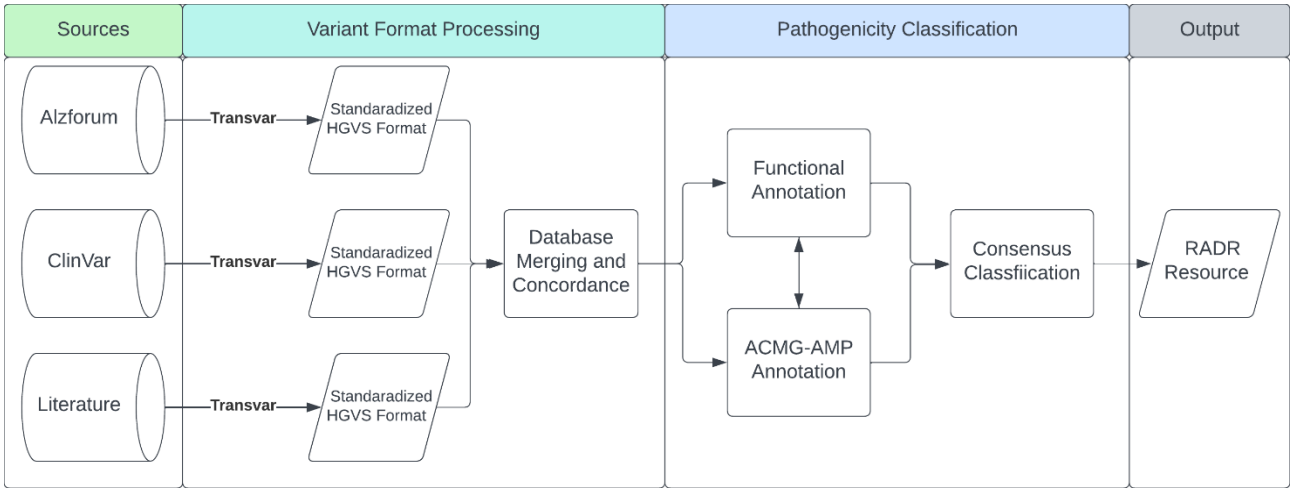


Figure 1a. RADR Resource Pipeline

Methods

Database sources

We commenced our database curation with Alzforum, a repository for expert-curated variants associated with Alzheimer's disease (AD) in known AD genes. This repository lists variants in six essential genes associated with AD, namely: Amyloid Precursor Protein (APP), Presenilin 1 (PSEN1), Presenilin 2 (PSEN2), Microtubule-Associated Protein Tau (MAPT), and Triggering Receptor Expressed on Myeloid cells 2 (TREM2), and APOE. To focus on the rare variants, our analysis at the point of curation included only the first five genes. Alzforum provides a list of reported variants in the literature, briefly describing associated clinical, and neuropathological features and functional impacts.[18] It also classifies the pathogenicity of variants of APP, PSEN1, and PSEN2 based on the American College of Medical Genetics and Genomics and Association of Molecular Pathology (ACMG-AMP) for the classification of genes that cause Mendelian disorders.[19] We extracted the variants and standardized them into HGVS format in both GrCH37 and GrCH38 builds using TransVar, a multi-way annotator for genetic elements and genetic variations.[20]

Next, we included variants in ClinVar, an open-access repository of relationships between human genetic variation and phenotypes, supported by evidential data. Despite its liberal classification of variants, ClinVar serves as the largest variant database. ClinVar accepts variants in any part of the genome and is interpreted for any type of condition.[21] Using keyword search with the terms "dementia" and "Alzheimer," we extracted relevant variants related to ADRD phenotypes, which we then converted to the standardized HGVS format via TransVar.

We additionally incorporated data from the Dominantly Inherited Alzheimer Network (DIAN) Observational Study, a valuable source of information on dominantly inherited Alzheimer's disease (DIAD) developed at Washington University at St. Louis. This study offers a catalog of pathogenic variants of the genes PSEN1, PSEN2, and APP, along with clinical risk factors as observed through clinical trials.

We conducted a literature review of variants associated with Alzheimer's Disease and Related dementia and added the variants along with the odds ratios associated with the study to RADR. We standardized the variants into HGVS format using our TransVar pipeline.

Functional and Clinical Interpretation Annotations

To make RADR a more comprehensive repository, we employed the ANNOVAR software to functionally annotate the variants. ANNOVAR is an annotation tool that is used to annotate variant exonic functions, identify variants in specific regions of the genome, as well as identify variants that are documented in specific databases.[22] Additionally, we extracted clinvar pathogenicity from ANNOVAR for consensus classification.

For clinical interpretation of genetic variants, we utilized InterVar, a bioinformatics software guided by the ACMG/AMP 2015 criteria. The tool takes in variants and interprets variants as "Pathogenic", "Likely Pathogenic", "Uncertain Significance", "Benign", and "Likely Benign" based on the guidelines.[23] The results from InterVar, along with the classifications from ClinVar's ClinSig, were used for additional data-driven and evidence-based variant consensus classification shown in Figure 1b.

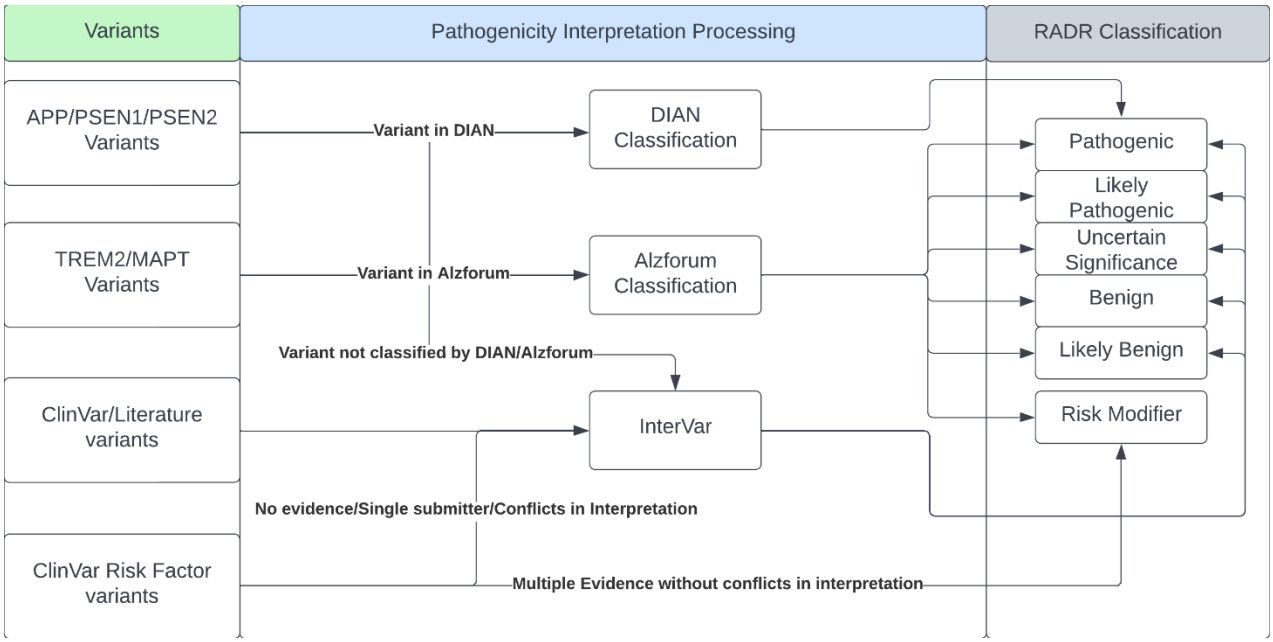


Figure 1b. Variant consensus classification flowchart

WES Variant Frequency Analysis

We utilized Mount Sinai Health System's BioMe Cohort for our whole exome sequencing (WES) data. Bio*Me* is a collection of 31,250 individual's de-identified DNA, plasma, clinical medical record and questionnaire data, and large-scale genome-wide genotype and exome-chip data. Quality control of genetic data was performed prior to analysis by the Institute of Personalized

Medicine (IPM) at Mount Sinai. The case-control derivation in BioMe used a curated list of patient visits, complete with International Classification of Diseases (ICD) codes and descriptions. Despite potential inaccuracies in the ICD codes, we augmented our precision by considering visit descriptions, medical history, and prescribed medications. We used a derivation of an internally validated dementia case-control algorithm to extract individuals in BioMe with ADRD phenotypes, including unspecified AD, late-onset AD (LOAD), early-onset AD (EOAD), frontotemporal dementia, vascular dementia, Lewy-body dementia, HIV dementia, and a family history of dementia. Additionally, we split unspecified dementia into AD Drugs + Unspecified dementia, and added another category “Dementia Drugs Taken” that applied to controls that take dementia drugs. The drugs we included as part of the case control process include: donepezil, galantamine, rivastigmine, memantine. The internally validated case-control algorithm utilizes ICD codes, description of visit, prescriptions and medical history to determine a patient’s phenotype classification.

As an additional WES dataset, we utilized the United Kingdom Biobank (UKB). UKB is a large-scale biomedical database and research resource, containing in-depth genetic and health information from half a million UK participants.[24] For our study, we analyzed genetic and phenotypic data of 201,136 individuals and case-control derivation was conducted to extract Alzheimer’s Disease (AD), Vascular dementia (V), Frontotemporal dementia (FTD), Lewy-body dementia (DLB), other dementia subtypes (O) and no dementia subtype specified (N). We determined these phenotypes by utilizing ICD Code, hospital records, death records and primary general practioner visit descripts with the process outlined by Wilkinson et al.[25]

After determining case-control status of patients, we extracted the carriers and non-carriers frequencies stratified by variant as well as phenotype. We conducted a Fisher’s Exact Test on the carrier and non-carrier frequencies of ADRD phenotypes and Control. Additionally we extracted the allele frequencies of pathogenic, likely pathogenic and risk factor/modifier variants within BioMe and UKB and compared it to Gnomad 3.12v allele frequencies. Multiple linear regression, and Cox regression were also conducted using the stats package and survival package on R.

Ancestry of each sample was derived from a self-report questionnaire and was also determined genetically using a combination of the 1000g reference panel with the BioMe GSA_GDA genotype dataset. The intersected data was mixed with ADMIXTURE and GrafPOP as evaluation and decision tool. (Cite Michael Preuss and BioMe team).We utilized BCFtools for manipulating whole-exome sequencing (WES) data from BioMe and UKB.[26] The vcfR package developed by Knuas et al. (2017) in R was applied to analyze WES data via VCF files.[27]All analyses were performed on the Minerva High-Performance Computing platform at Mount Sinai.

Results

Developing the RADR Database for ADRD Rare Variants

The juxtaposition of data from Alzforum, ClinVar, and DIAN underscores the critical challenge of fragmentation in Alzheimer's disease-related dementias (ADRD) genetic resources. As depicted

in Figure 2, a comparative analysis was performed between the APP, PSEN1, and PSEN2 variants in Alzforum and those in DIAN and ClinVar, respectively. A significant overlap between Alzforum and DIAN is evident (n=121) across the APP, PSEN1, and PSEN2 genes. However, DIAN encompasses certain variants absent in Alzforum. This pattern intensifies when comparing the genetic variants in Alzforum with those in ClinVar. Notwithstanding the larger overlap (n=80) across APP, PSEN1, PSEN2, ClinVar encompasses a substantial number of additional variants (n=597) absent in Alzforum and DIAN.

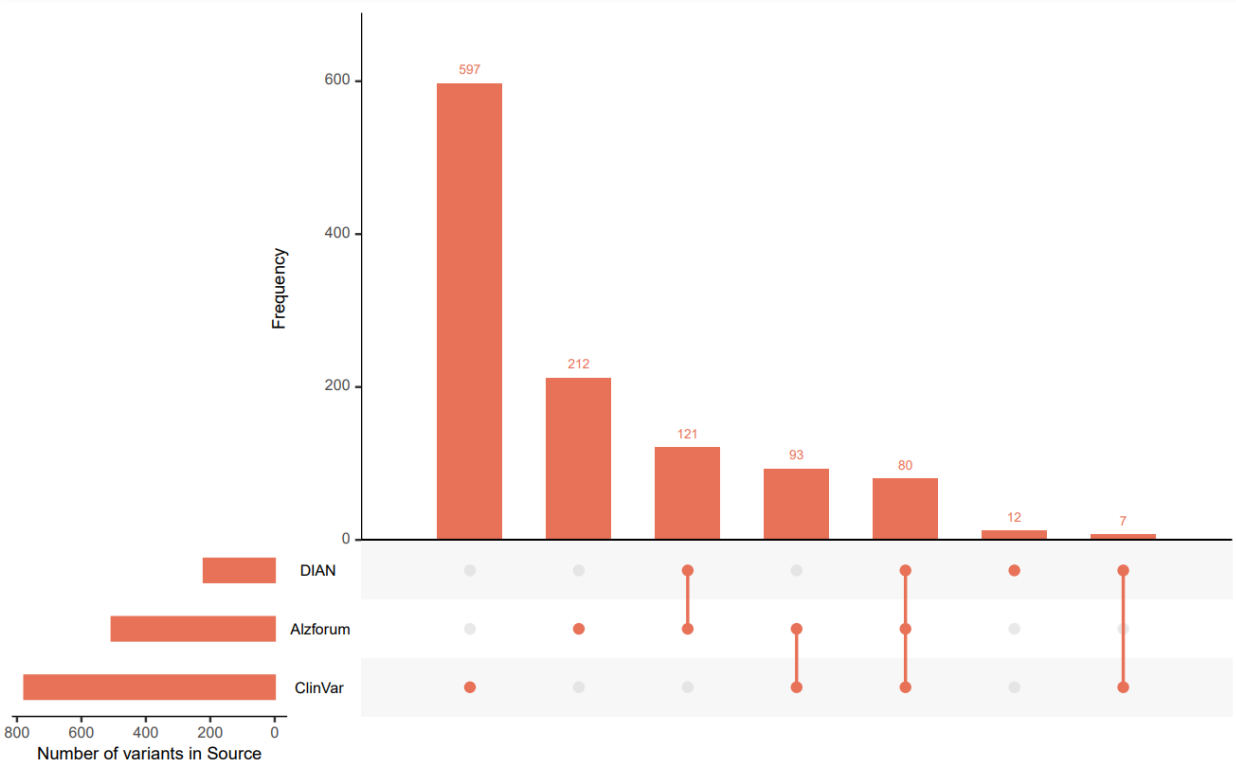


Figure 2. Comparative analysis of APP/PSEN1/PSEN2 variants in ClinVar, Alzforum, and DIAN.

Following the integration of variant data into the RADR database, we examined the distribution of genetic variants stratified by pathogenicity, as determined by our consensus classification algorithm. Figure 3a portrays the genes within RADR that possess over ten pathogenic, likely pathogenic, and risk factor/modifier genetic variants. PSEN1 and SQSTM1 genes are associated with a higher number of ADRD phenotype variants. Notably, the majority of PSEN1 mutations fall under the categories of pathogenic and likely pathogenic. As anticipated, many of the variants across the genes with pathogenic, likely pathogenic, and risk factor/modifier variants are deemed variants of uncertain significance.

For a detailed gene distribution analysis stratified by pathogenicity, please refer to Supplemental Figure 1.

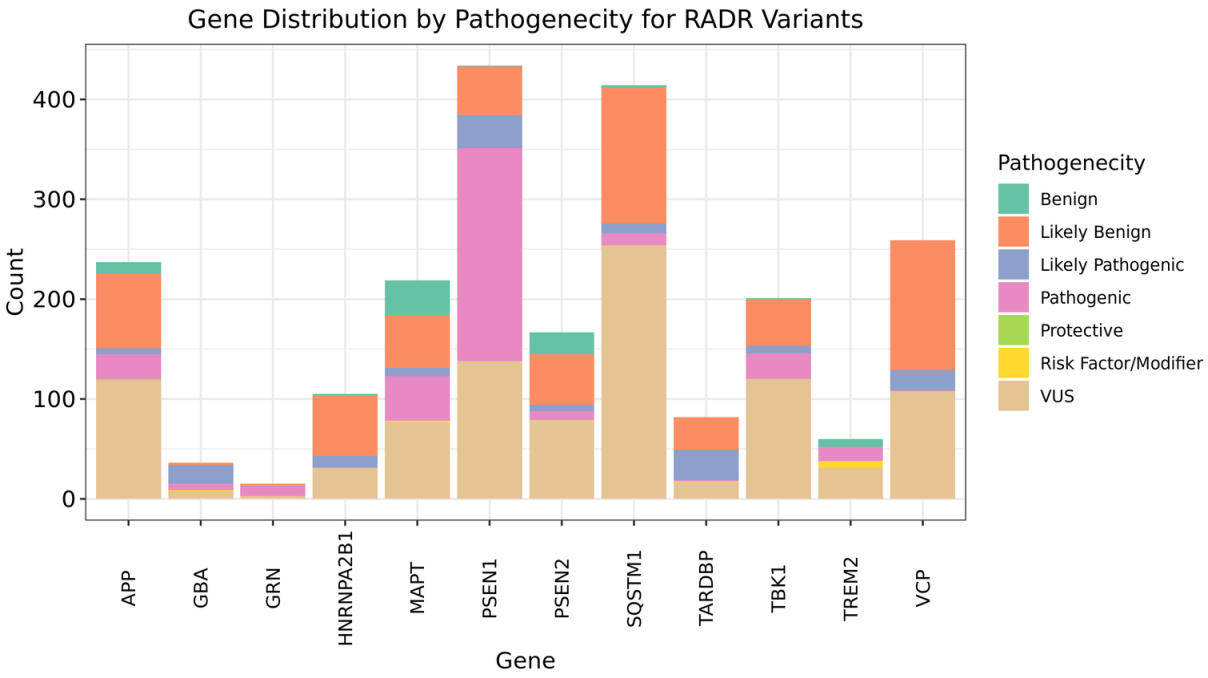


Figure 3a. Gene Distribution of Pathogenicity for pathogenic, likely pathogenic, and risk factor/modifier RADR variants.

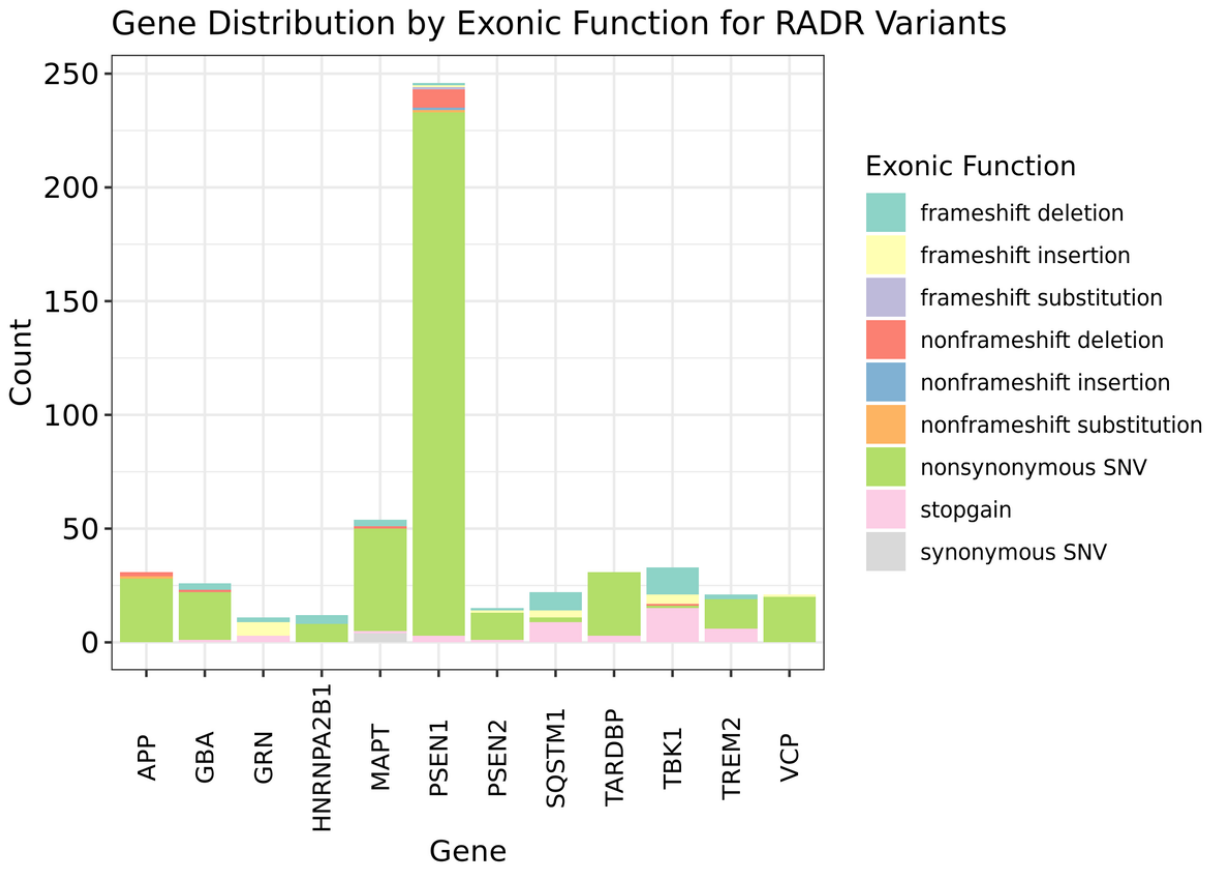


Figure 3b. Gene distribution of genes with pathogenic, likely pathogenic, and risk factor/modifier variants stratified by exonic function.

Next, we investigated the genetic distribution of pathogenic, likely pathogenic, and risk factor/modifier variants, stratified by exonic function. The majority of these variants display a form of nonsynonymous variation. Interestingly, almost all curated genes, excluding those of GRN, SQSTM1, and TBK1 genes, have the majority of their pathogenic, likely pathogenic, and risk factor/modifier variants classified as nonsynonymous single nucleotide variants (SNVs). For a comprehensive breakdown of the genetic distribution of variants stratified by exonic function, please refer to Supplemental Figure 2.

Application of RADR to Mount Sinai's BioMe Biobank

Leveraging the RADR database, we conducted an analysis of carrier rates for rare pathogenic, likely pathogenic, and risk factor/modifier variants within the BioMe Biobank. Variants classified under these categories are presumed to significantly influence phenotype and are consequently subject to intense selective pressure, hence their rarity. Figure 4a and Figure 4b illustrate the phenotype-specific carrier frequency of pathogenic, likely pathogenic, and risk factor/modifier variants respectively. Evidently, these variants within the BioMe Biobank occur in both distinct cases and controls, suggesting incomplete penetrance.

Interestingly, the HFE:p.Cys287Tyr variant is observed in multiple instances of both individuals with ADRD phenotypes and control cases. This variant has been reported in ClinVar by multiple submitters to be either pathogenic or a risk factor. The HFE gene produces a protein that detects iron levels in the body and defects in this gene are mainly associated with Hemochromatosis, but also linked to AD.[30] Additionally, the PSEN1:p.Gly206Ala variant warrants attention as it has been identified as a founder variant in the Caribbean Hispanic population.

The GBA:p.Asn409Ser variant is also seen in the biobank population, and is recognized in literature as pathogenic for Parkinson's disease and Lewy Body Dementia, with a prevalence in the Ashkenazi Jewish population.[31] Despite the limited number of patients in BioMe (n=4), further ancestry analysis is required to confirm this population-specificity.

Among risk factor variants , TREM2:p.Leu211Pro showed the highest carrier frequency in BioMe. This variant, categorized as benign or likely benign, is not linked to any ADRD phenotypes in ClinVar. However, Alzforum cites it as a potential risk factor/modifier for AD and FTD, due to its possible correlation with AD in African Americans.[32] Furthermore, TREM2:p.Arg47His was observed in patients with AD, unspecified dementia, and a family history of dementia. Two independent 2013 studies associated this missense variant with late-onset Alzheimer's disease.[33]

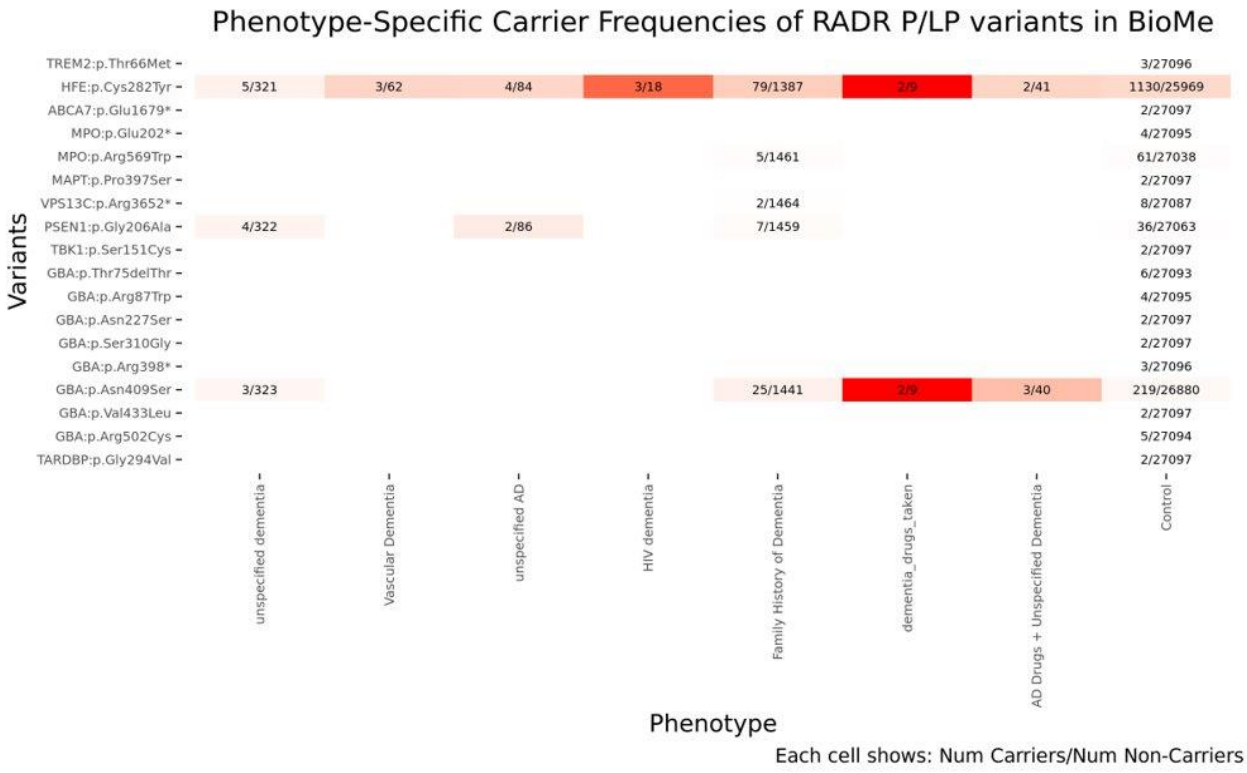


Figure 4a. Heatmap showing phenotype specific carrier frequency of RADR pathogenic and likely pathogenic variants found in the BioMe Biobank. This figure is a subset of all pathogenic and likely pathogenic variants found in BioMe. A complete heatmap is found in Supplemental Figure 3.

*henotype	*arriers	*lon-Carriers	*isher Test FDR	*dds Ratio
AD Drugs + Unspecified Dementia		7	.03*	.79
ontrol	.489	5610	.00	.00
ementia Drugs Taken	.	'	.05>*	.83
arly Onset Alzheimer's Disease		'	.08	.91
amily History of Dementia	.19	.347	.00*	.52
rontotemporal Dementia		.	.00	.00
IV Dementia		.8	.11	.87
luntington's Disease Dementia	.		.05	/A
ewy Body Dementia	.		.33	.87
ate Onset Alzheimer's Disease		9	.27	.00
arkinson Dementia	.		.43	.91
enile Dementia	.	4	.00	.51
nspecified AD	'	1	.34	.49
nspecified Dementia	.3	13	.27	.71
ascular Dementia		2	.00	.83

Table 1a. Fisher Test and Odds Ratio of phenotypes with carriers of pathogenic and likely pathogenic variants in BioMe. * indicates statistically significant p-values at an α = 0.05.

Phenotype ~ genotype + covariates (group all ADRD phenotypes for now)

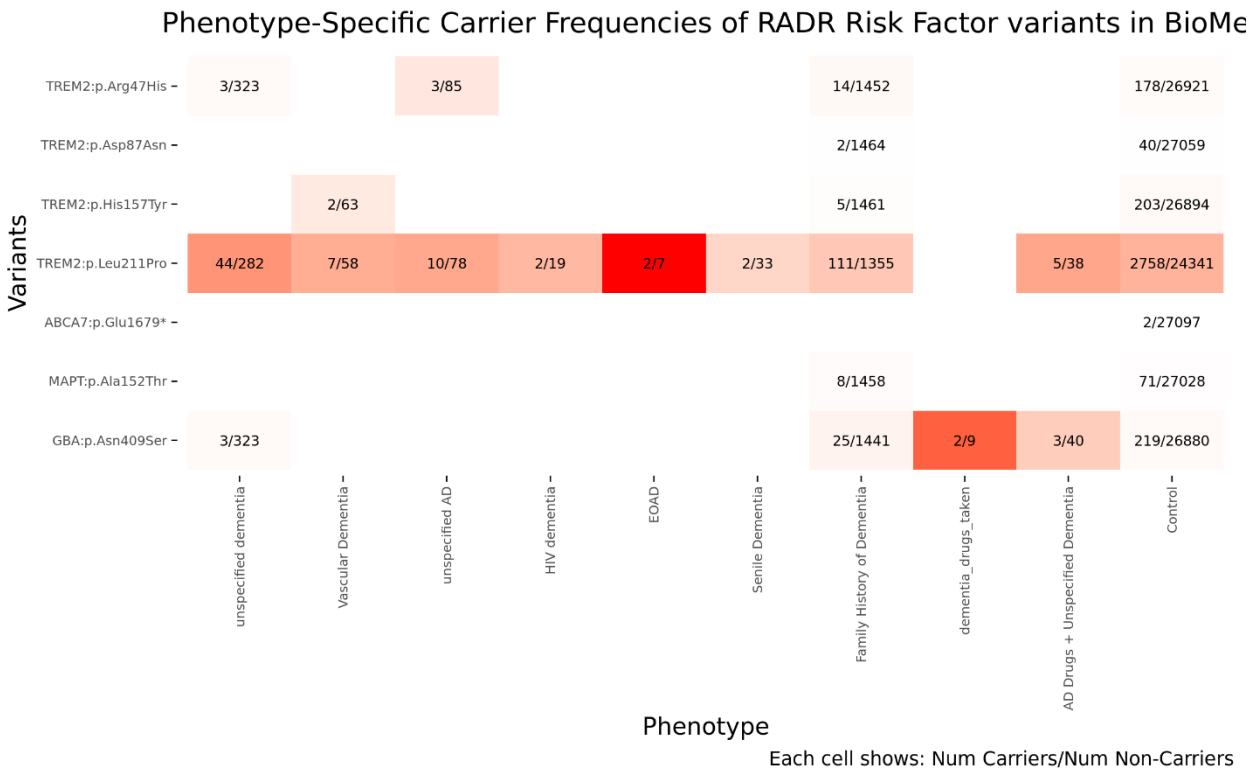


Figure 4b. Heatmap showing phenotype specific carrier frequency of RADR risk factor/modifier variants found in the BioMe Biobank.

Phenotypes	*Carriers*	*Non-Carriers*	*Fisher Test P-Value*	*Odds Ratio*
AD Drugs + Unspecified Dementia	9	34	0.11	1.82
Control	3441	23658	1.00	1.00
Dementia Drugs Taken	2	9	0.64	1.53
Early Onset Alzheimer's Disease	3	6	0.10	3.44
Family History of Dementia	164	1302	0.09	0.87
Frontotemporal Dementia	0	4	1.00	0.00
HIV Dementia	2	19	1.00	0.72
Huntington's Disease Dementia	0	1	1.00	0.00
Lewy Body Dementia	2	5	0.22	2.75
Late Onset Alzheimer's Disease	3	36	0.47	0.57
Parkinson Dementia	1	9	1.00	0.76
Senile Dementia	3	32	0.62	0.64
Unspecified AD	15	73	0.26	1.41
Unspecified Dementia	50	276	0.16	1.25
Vascular Dementia	9	56	0.71	1.10

Table 1b. Fisher Test and Odds Ratio of phenotypes with carriers of risk factor/modifier variants in BioMe.

Following the initial carrier frequency analysis, we applied a Fisher's Exact Test to compare the number of carriers and non-carriers of pathogenic and likely pathogenic variants in RADR for ADRD phenotypes and controls, as presented in Table 1a. The phenotypes that resulted in a statistically significant p-value at an α of 0.05 were patients classified as Dementia Drugs Taken, patients with a family history of dementia, and patients classified as AD Drugs + unspecified dementia. Further downstream analysis needs to be conducted to elucidate a possible reason for this result. A similar test was performed on the risk factor/modifier variants in BioMe, which failed to yield any significant results for any phenotypes when compared to control. One major factor limiting this analysis is the lack statistical power, which can be remediated by finding more pathogenic variants or analyzing more genomes of patients with ADRD phenotypes.

Validation of RADR resource in UKB

Our next step was to validate the RADR resource through its application to the UK Biobank (UKB). Mirroring our methodology used in the BioMe variant analysis, we conducted a preliminary evaluation of the occurrence frequencies of rare pathogenic, likely pathogenic, and risk factors/modifiers curated in RADR. Figure 5a and Figure 5b graphically represent the phenotype-specific carrier frequency of these categories of variants. Variants such as HFE:chr6:g.26093233G>A and SNCA:p.Gly735Ter are observed in both specific cases and controls, implying incomplete penetrance.

HFE:chr6:g.26093233G>A is notably enriched in AD, No dementia subtype specifieed, and control cases and is characterized as pathogenic in AD according to ClinVar. This variant resides in a intronic spliced donor site, and it is anticipated to impact mRNA splicing, potentially resulting in a substantially modified protein.[34]

Additionally, SNCA:p.Gly73Ser demonstrates enrichment in Vascular Dementia, Alzheimer's Disease, No dementia subtype specified, and control cases. The synuclein gene (SNCA) is responsible for producing alpha-synuclein, a protein considered to play a role in presynaptic signaling and implicated in Parkinson's disease pathogenesis. Despite this, SNCA peptides constitute a major component of amyloid plaques observed in AD. This particular variant is linked to Lewy Body Dementia in ClinVar.[35]

For risk factors of RADR, TREM2:p.Asp87Asn was the most enriched in UKB. While this variant is not associated with any ADRD phenotypes according to ClinVar, Alzforum classifies it as a potential risk factor/modifier, attributing to two studies correlating the variant with AD.[33] Nonetheless, it's crucial to note that subsequent replication studies have not confirmed this association, thus necessitating further analysis to definitively establish ADRD association.

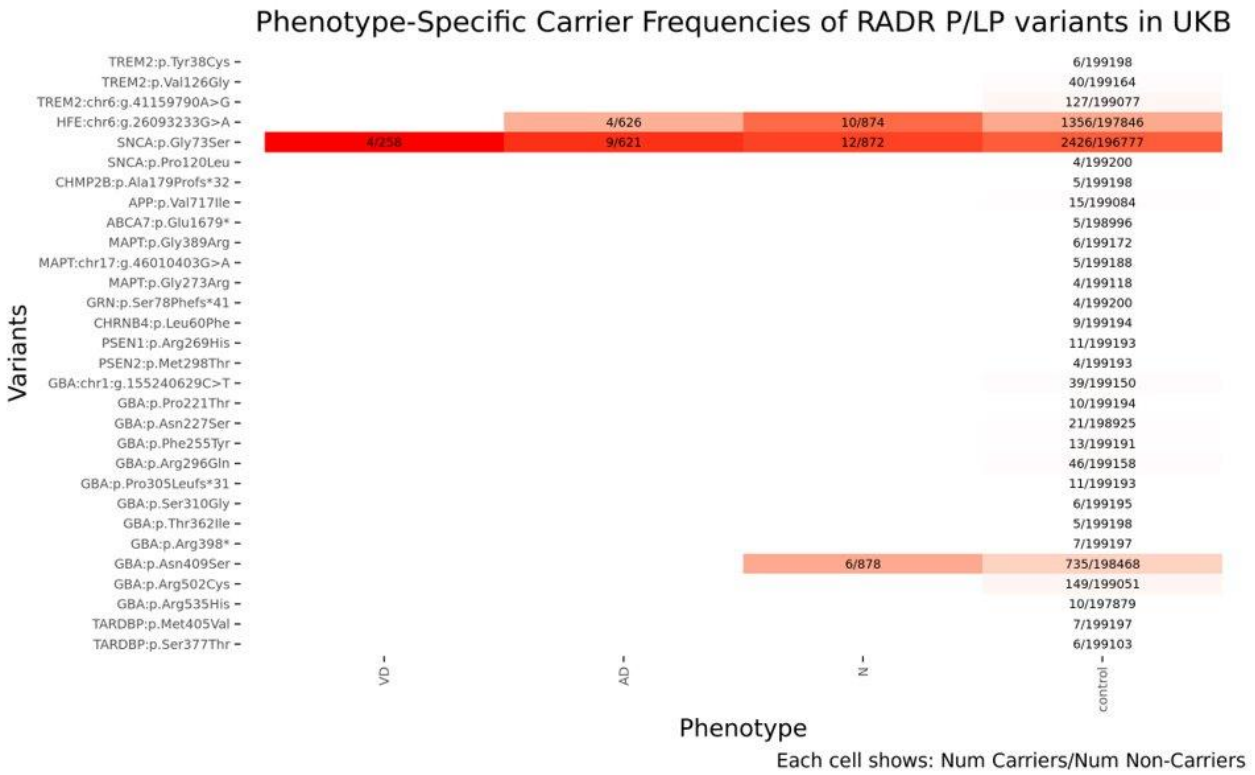


Figure 5a. Heatmap showing phenotype-specific carrier frequency of RADR pathogenic and likely pathogenic variants with more than 2 carriers found in UKB. This figure is a subset of all pathogenic and likely pathogenic variants found in UKB. A complete heatmap is found in Supplemental Figure 4.

Phenotype	*Carriers*	*Non-Carriers*	*Fisher Test P-Value*	*Odds Ratio*
Alzheimer's Disease	21	609	0.26	1.29
Control	5170	194033	1.00	1.00
Frontotemporal Dementia	5	257	0.69	0.73
Lewy Body Dementia	1	2	0.08	18.77
No dementia subtype specified	1	49	1.00	0.77
Another dementia subtype	33	851	0.14	1.46
Vascular Dementia	3	100	0.75	1.13

Table 2a. Fisher Test and Odds Ratio of phenotypes with carriers of pathogenic and likely pathogenic variants in UKB.

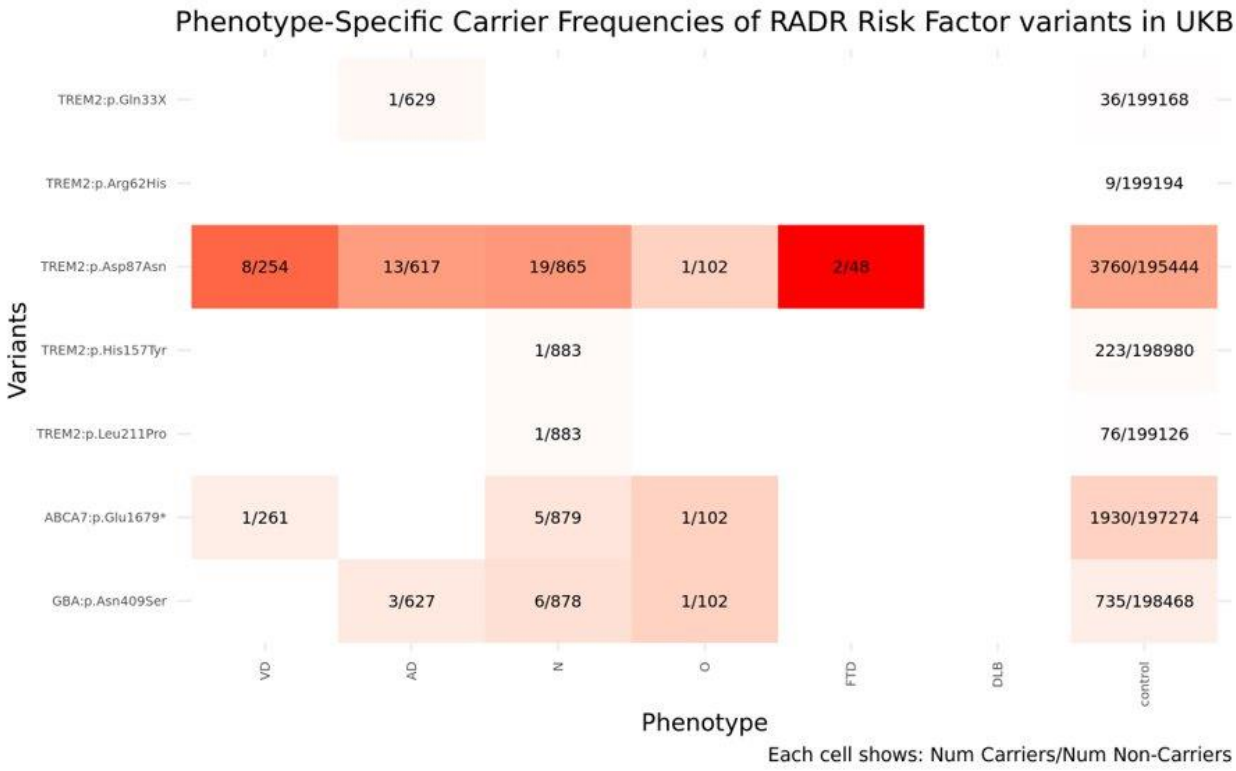


Figure 5b. Heatmap showing phenotype specific carrier frequency of RADR risk factor/modifier variants in UKB. This figure is a subset of all pathogenic and likely pathogenic variants found in UKB.

Phenotypes	*Carriers*	*Non-Carriers*	*Fisher Test P-Value*	*Odds Ratio*
Alzheimer's Disease	17	613	0.44	0.79
Control	6748	192456	1.00	1.00
Frontotemporal Dementia	9	253	0.86	1.01
Lewy Body Dementia	0	3	1.00	0.00
No dementia subtype specified	2	48	0.69	1.19
Other dementia subtype	32	852	0.71	1.07
Vascular Dementia	3	100	1.00	0.86

Table 2b. Fisher Test and Odds Ratio of phenotypes with carriers of risk factor/modifier variants in UKB.

In parallel with the Fisher's Exact Test conducted for ADRD phenotypes in BioMe, we applied a similar analysis to the carriers of rare pathogenic, likely pathogenic, and risk factor/modifier RADR variants, as depicted in Tables 2a and 2b. Despite the increased number of ADRD phenotype cases in comparison to BioMe, we found no significant associations for carriers of pathogenic, likely pathogenic, and risk factor/modifier variants.

Evaluation of P/LP/Risk Factor Frequencies in BioMe and UKB

Building on our analysis of the carrier frequencies for pathogenic, likely pathogenic, and risk factor variants within BioMe and UKB, we subsequently examined the allele frequencies of these variants, as visualized in Figure 6 and documented in Table 3. Owing to the selective pressure exerted on pathogenic and likely pathogenic variants, we anticipated lower frequencies for these

variants compared to risk factors. Our data from both BioMe and UKB, presented in Figure 6, supports this expectation. Most pathogenic and likely pathogenic variants present a minor allele frequency (MAF) in the range of 0-0.1%, although certain variants display greater enrichment.

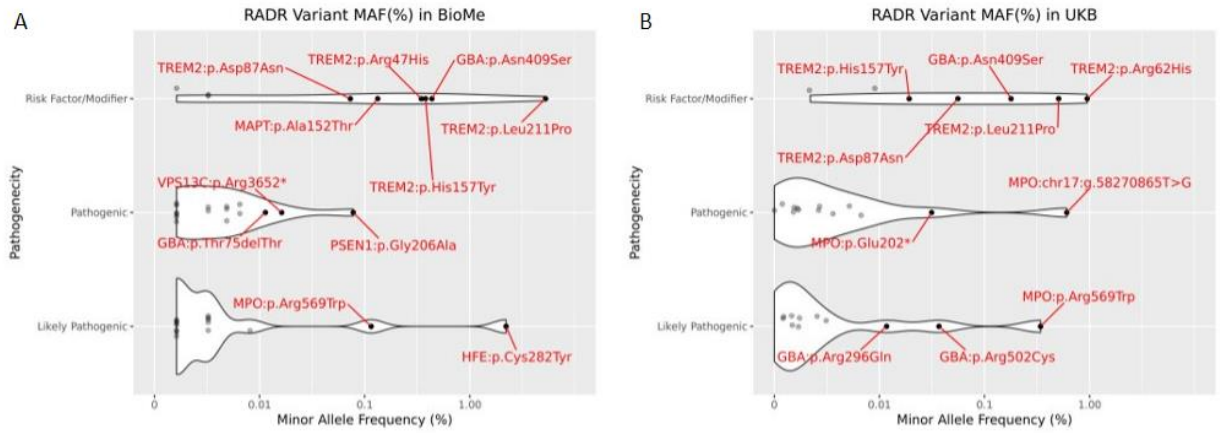


Figure 6. Minor Allele Frequencies of RADR P/LP/Risk Factor/Modifiers variants in BioMe and UKB. Panel A is a violin plot of MAF in BioMe. Panel B is a violin plot of MAF in UKB. Variants are highlighted that have a MAF > 0.01%.

Variant	*Pathogenicity*	*BioMe AF*	*UKB AF*	*Gnomad 3.12 Allele Frequency*
ABCA7:p.Arg1561*	Pathogenic	1.62269e-05	1.5e-05	1.971e-05
ABCA7:p.Glu1679*	Risk Factor/Modifier	3.24538e-05	9.2e-05	3.945e-05
GBA:p.Arg398*	Pathogenic	4.86808e-05	1.7e-05	.
GBA:p.Arg502Cys	Likely Pathogenic	8.11346e-05	0.000369	7.28e-05
GBA:p.Arg87Trp	Pathogenic	6.49077e-05	7e-06	1.972e-05
GBA:p.Asn227Ser	Pathogenic	4.86808e-05	5.2e-05	9.861e-05
GBA:p.Asn409Ser	Risk Factor/Modifier	0.00436504	0.001789	0.0018
GBA:p.Leu199Aspfs*62	Pathogenic	1.62269e-05	2e-06	.
GBA:p.Phe252Ile	Likely Pathogenic	1.62269e-05	5e-06	6.572e-06
GBA:p.Ser235Pro	Likely Pathogenic	1.62269e-05	2e-06	6.596e-06
GBA:p.Ser310Gly	Likely Pathogenic	3.24538e-05	1.5e-05	3.942e-05
GBA:p.Val433Leu	Likely Pathogenic	3.24538e-05	5e-06	6.569e-06
MPO:p.Arg569Trp	Likely Pathogenic	0.00115211	0.003414	0.0017
MPO:p.Glu202*	Pathogenic	6.49077e-05	0.000314	0.0001
PSEN1:p.Ala79Val	Pathogenic	1.62269e-05	5e-06	3.286e-05
PSEN1:p.Val412Ile	Pathogenic	3.24538e-05	5e-06	1.972e-05
TBK1:p.Glu643delGlu	Likely Pathogenic	1.62311e-05	1.2e-05	.
TREM2:p.Asp87Asn	Risk Factor/Modifier	0.000730211	0.000558	0.0014
TREM2:p.Gln33X	Risk Factor/Modifier	3.24538e-05	2.2e-05	1.971e-05
TREM2:p.His157Tyr	Risk Factor/Modifier	0.00379735	0.000192	0.0019
TREM2:p.Leu211Pro	Risk Factor/Modifier	0.0526888	0.005076	0.0386

TREM2:p.Thr66Met	Pathogenic	4.86808e-05	2.7e-05	5.912e-05
TREM2:p.Tyr38Cys	Pathogenic	1.62269e-05	2e-06	.
VCP:p.Pro137Ser	Likely Pathogenic	1.62269e-05	2e-06	.

Table 3. Allele Frequencies of P/LP/Risk Factors present in both UKB and BioMe compared to gnomAD. Variants with "." in Gnomad 3.12 Allele Frequency column have not been recorded in Gnomad.

Table 3 shows the allele frequencies of the RADR variants identified in both BioMe and UKB in conjunction with those reported in gnomAD. Although most variants display consistent frequencies across both BioMe and UKB, specific variants, such as GBA:p.Asn409Ser and TREM2:p.His157Tyr, show higher frequencies in BioMe and UKB. This discrepancy can potentially be attributed to several factors.

The differing population structures of the three databases could account for the observed variation. BioMe boasts a more diverse biobank compared to the predominantly white-Caucasian population represented in UKB. Consequently, certain variants that are particularly enriched in specific populations may be more prevalent in BioMe. To reconcile these disparities in frequency, further ancestral analysis is warranted.

A comprehensive list of all pathogenic, likely pathogenic, and risk factor/modifier variants for BioMe and UKB can be accessed in Supplemental Tables 1 and 2, respectively.

Ancestry-Specific Variants that Contribute to ADRD

Our analysis of each variant within BioMe also yielded variants that were enriched within certain ancestries. Of note, MAPT:Ala152Thr, a risk modifier variant, was enriched in hispanic population. PSEN1:Gly206Ala, a pathogenic variant, was enriched in the hispanic population as well in BioMe. This follows previous analysis of this variant, which has been linked to Alzheimer's disease in over 70 families of Carribean- Hispanic descent. (Lee et al. 2014). TREM2:Arg47His is another risk modifier variant which has been shown to increase risk for late-onset Alzheimer's Disease. While this variant was enriched in Population-specific frequency (BioMe & gnomAD)

Variant Penetrance

Familial cases (maybe a paragraph)

Discussion

This study highlights the importance of a comprehensive and unifying resource for understanding the genetic underpinnings of Alzheimer's disease and related dementias (ADRD). By developing RADR, we have made significant strides in harmonizing and standardizing the scattered variants associated with ADRD from multiple databases, including ClinVar, Alzforum, and DIAN.

One striking observation from our findings was the difference in variant classification between ClinVar, Alzforum, and the literature of ADRD WES studies, which underscores the necessity for a standardization pipeline and cohesive database. This discrepancy is indicative of the substantial fragmentation and incongruity within the field, reinforcing the need for a unified platform such as RADR, which aims to rectify these discrepancies and streamline data interpretation.

Looking ahead, there are several promising avenues for future research. One key direction involves analyzing the ancestry breakdown of variant frequencies in the BioMe cohort. As our results have suggested, population structure and ancestry play a crucial role in variant frequencies. More comprehensive and detailed analysis will help account for these variations, adding to the precision and reliability of our database.

Another compelling direction involves identifying familial cases from population cohorts. With the advent of large-scale biobanks and genomic databases, we have an unprecedented opportunity to detect familial aggregation of ADRD, which could provide valuable insights into both the heritable and sporadic forms of the disease. Furthermore, it can also help us determine the penetrance of variants and provide more clarity.

Validation of RADR in other large-scale sequencing datasets, such as the "All of Us" Research Program, will also be crucial for expanding the generalizability and robustness of the database. The “All of US” research program initiated by the National Institutes of Health (NIH) to gather health data from one million or more people in the United States.[36] The genetic and health data will be invaluable to identifying new variants as well as validating the association of establised variants.This approach will enable us to cross-validate our findings and ensure the reliability of RADR across different populations and cohorts.

Given the growing evidence for the role of rare variants in ADRD, the development of a rare variant Polygenic Risk Score (PRS) could prove instrumental in deciphering the genetic complexity of the disease. In addition, a combined PRS approach, encompassing both common and rare variants, could potentially fill the 'missing heritability' gap of ADRD and significantly enhance our predictive and diagnostic capabilities. Attempts have been made to combine Common and Rare PRS in the context of diabetes and prostate cancer, however, a standardized methodology for ADRD remains elusive.[37,38]

Finally, it's important to highlight that the development of RADR is an ongoing endeavor. As our understanding of the genetic architecture of ADRD continues to evolve, so will RADR. We envision RADR as a dynamic and continually updated repository, growing in tandem with the evolving nature of ADRD genetics research. As new genetic insights emerge and novel variants are identified, these will be incorporated into RADR, thus maintaining its relevance and value as a critical resource for researchers and clinicians alike. The newly announced Mount Sinai Million Health Discoveries program, an extension to the BioMe biobank as well as the “All of Us” reseach program are crucial to maintaining and applying RADR.[39]

Our work with RADR has only just begun, but the promise it holds is considerable. By providing a comprehensive, harmonized, and continuously updated resource, we hope to accelerate advances in our understanding of ADRD genetics and ultimately contribute to the development of targeted interventions and treatments for ADRD diseases.

Acknowledgements

The authors wish to acknowledge the BioMe and UKB biobanks and their participating patients and families that generously contributed the data. The authors would also wish to acknowledge all genetics research groups that have aided in mapping of ADRD genetics thus far. This work was supported by NIGMS R35GM138113 and ISMMS funds to KH. Large language models (LLM), including those provided through GPT models, is used in the initial drafts of coding, literature review, writing of this work. All final codes and texts have been extensively edited and verified by the authors.

I would like to thank my PI Dr. Kuan-lin Huang for his invaluable assistance and guidance with the development of RADR. I would also like to thank the members of the Computationalomics Laboratory and the Department of Genomics and Genetics at Mount Sinai.

COMPETING FINANCIAL INTERESTS

The authors declare no competing interests.

References

1. Manly JJ, Jones RN, Langa KM, Ryan LH, Levine DA, McCammon R, et al. Estimating the Prevalence of Dementia and Mild Cognitive Impairment in the US: The 2016 Health and Retirement Study Harmonized Cognitive Assessment Protocol Project. JAMA Neurol 2022;79(12):1242-9.
2. Andrews SJ, Renton AE, Fulton-Howard B, Podlesny-Drabiniok A, Marcora E, Goate AM. The complex genetic architecture of Alzheimer's disease: novel insights and future directions. EBioMedicine. 2023;90:104511. doi:10.1016/j.ebiom.2023.104511
3. . Loy CT, Schofield PR, Turner AM, Kwok JBJ. Genetics of dementia. Lancet 2014;383:828-40.
4. Holtzman DM, Herz J, Bu G. Apolipoprotein E and apolipoprotein E receptors: Normal biology and roles in Alzheimer disease. Cold Spring Harb Perspect Med 2012;2(3):a006312.
5. Michaelson DM. APOE ε4: The most prevalent yet understudied risk factor for Alzheimer’s disease. Alzheimers Dement 2014;10:861-8.
6. Bellenguez C., Küçükali F., Jansen I.E., et al. New insights into the genetic etiology of Alzheimer's disease and related dementias. Nat Genet. 2022;54:412–436.

7. Wightman D.P., Jansen I.E., Savage J.E., et al. A genome-wide association study with 1,126,563 individuals identifies new risk loci for Alzheimer's disease. Nat Genet. 2021;53:1276–1282.
8. Holstege, H. et al. Exome sequencing identifies rare damaging variants in ATP8B4 and ABCA1 as risk factors for Alzheimer's disease. Nat. Genet. 54, 1786–1794 (2022).
9. Sims, R. et al. Rare coding variants in PLCG2, ABI3, and TREM2 implicate microglial-mediated innate immunity in Alzheimer's disease. Nat. Genet. 49, 1373–1384 (2017).
10. Jonsson, T. et al. Variant of TREM2 associated with the risk of Alzheimer's disease. N. Engl. J. Med. 368, 107–116 (2013).
11. Guerreiro, R. et al. TREM2 variants in Alzheimer's disease. N. Engl. J. Med. 368, 117–127 (2013).
12. Vardarajan, B. N. et al. Coding mutations in SORL1 and Alzheimer disease. Ann. Neurol. 77, 215–227 (2015).
13. Vardarajan, B. N. et al. Rare coding mutations identified by sequencing of Alzheimer disease genome-wide association studies loci. Ann. Neurol. 78, 487–498 (2015).
14. Steinberg, S. et al. Loss-of-function variants in ABCA7 confer risk of Alzheimer's disease. Nat. Genet. 47, 445–447 (2015).
15. Logue, M. W. et al. Two rare AKAP9 variants are associated with Alzheimer's disease in African Americans. Alzheimers Dement. J. Alzheimers Assoc. 10, 609-618.e11 (2014).
16. Jun, G. et al. PLXNA4 is associated with Alzheimer disease and modulates tau phosphorylation. Ann. Neurol. 76, 379–392 (2014).
17. Wetzel-Smith, M. et al. A rare mutation in UNC5C predisposes to Alzheimer's disease and increases neuronal cell death. Nat. Med. 20, 1452–1457 (2014).
18. http://www.alzforum.org/mutations. 4/21/2023.
19. Richards, S., Aziz, N., Bale, 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 17, 405–423 (2015). https://doi.org/10.1038/gim.2015.30
20. Zhou, W., Chen, T., Chong, Z., et al., TransVar: a multilevel variant annotator for precision genomics, Nature Methods 12 p1002 (2015). https://doi.org/10.1038/nmeth.3622
21. Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, Gu B, Hart J, Hoffman D, Jang W, Karapetyan K, Katz K, Liu C, Maddipatla Z, Malheiro A, McDaniel K, Ovetsky M, Riley G, Zhou G, Holmes JB, Kattman BL, Maglott DR. ClinVar: improving access to variant interpretations and supporting evidence. Nucleic Acids Res . 2018 Jan 4. PubMed PMID: 29165669 .
22. Kai Wang, Mingyao Li, Hakon Hakonarson, ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data, Nucleic Acids Research, Volume 38, Issue 16, 1 September 2010, Page e164, https://doi.org/10.1093/nar/gkq603
23. Quan Li and Kai Wang. InterVar: Clinical interpretation of genetic variants by ACMG-AMP 2015 guideline(The American Journal of Human Genetics 100, 1-14, February 2, 2017,http://dx.doi.org/10.1016/j.ajhg.2017.01.004)
24. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. (2015) UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. PLoS Med 12(3): e1001779. https://doi.org/10.1371/journal.pmed.1001779

25. Wilkinson T, Schnier C, Bush K, et al. Identifying dementia outcomes in UK Biobank: a validation study of primary care, hospital admissions and mortality data. Eur J Epidemiol. 2019;34(6):557-565. doi:10.1007/s10654-019-00499-1
26. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H. Twelve years of SAMtools and BCFtools. Gigascience. 2021 Feb 16;10(2):giab008. doi: 10.1093/gigascience/giab008. PMID: 33590861; PMCID: PMC7931819.
27. Knaus, Brian J., and Niklaus J. Grunwald. 2017. VCFR: a package to manipulate and visualize variant call format data in R. Molecular Ecology Resources 17(1):44-53. http://dx.doi.org/10.1111/1755-0998.12549.
28. HFE variant
29. Athan ES, Williamson J, Ciappa A, et al. A founder mutation in presenilin 1 causing early-onset Alzheimer disease in unrelated Caribbean Hispanic families. JAMA. 2001;286(18):2257-2263. doi:10.1001/jama.286.18.2257
30. Orme T, Hernandez D, Ross OA, et al. Analysis of neurodegenerative disease-causing genes in dementia with Lewy bodies. Acta Neuropathol Commun. 2020;8(1):5. Published 2020 Jan 29. doi:10.1186/s40478-020-0879-z
31. Ruskey JA, Greenbaum L, Roncière L, et al. Increased yield of full GBA sequencing in Ashkenazi Jews with Parkinson's disease. Eur J Med Genet. 2019;62(1):65-69. doi:10.1016/j.ejmg.2018.05.005
32. Jin SC, Carrasquillo MM, Benitez BA, Skorupa T, Carrell D, Patel D, Lincoln S, Krishnan S, Kachadoorian M, Reitz C, Mayeux R, Wingo TS, Lah JJ, Levey AI, Murrell J, Hendrie H, Foroud T, Graff-Radford NR, Goate AM, Cruchaga C, Ertekin-Taner N. TREM2 is associated with increased risk for Alzheimer's disease in African Americans. Mol Neurodegener. 2015 Apr 10;10:19. PubMed.
33. Guerreiro R, Wojtas A, Bras J, Carrasquillo M, Rogaeva E, Majounie E, Cruchaga C, Sassi C, Kauwe JS, Younkin S, Hazrati L, Collinge J, Pocock J, Lashley T, Williams J, Lambert JC, Amouyel P, Goate A, Rademakers R, Morgan K, Powell J, St George-Hyslop P, Singleton A, Hardy J, Alzheimer Genetic Analysis Group. TREM2 variants in Alzheimer's disease. N Engl J Med. 2013 Jan 10;368(2):117-27. Epub 2012 Nov 14 PubMed.
34. National Center for Biotechnology Information. ClinVar; [VCV002163003.1], https://www.ncbi.nlm.nih.gov/clinvar/variation/VCV002163003.1 (accessed May 31, 2023).
35. National Center for Biotechnology Information. ClinVar; [VCV001065637.9], https://www.ncbi.nlm.nih.gov/clinvar/variation/VCV001065637.9 (accessed May 31, 2023).
36. All of US research program overview - NIH. (2021, July 16). https://allofus.nih.gov/about/program-overview
37. Darst BF, Sheng X, Eeles RA, Kote-Jarai Z, Conti DV, Haiman CA. Combined Effect of a Polygenic Risk Score and Rare Genetic Variants on Prostate Cancer Risk. Eur Urol. 2021;80(2):134-138. doi:10.1016/j.eururo.2021.04.013
38. Dornbos, P., Koesterer, R., Ruttenburg, A. et al. A combined polygenic score of 21,293 rare and 22 common variants improves diabetes diagnosis based on hemoglobin A1C levels. Nat Genet 54, 1609–1614 (2022). https://doi.org/10.1038/s41588-022-01200-1

39. Mount Sinai Million Health Discoveries Program: Icahn School of Medicine. Icahn School of Medicine at Mount Sinai. (n.d.). https://icahn.mssm.edu/research/ipm/programs/mount-sinai-million
40. Jake R Conway and others, UpSetR: an R package for the visualization of intersecting sets and their properties, Bioinformatics, Volume 33, Issue 18, September 2017, Pages 2938–2940, https://doi.org/10.1093/bioinformatics/btx364

Supplemental Information

ariant	athogenicity	llele Frequency	nomad 3.12 Allele Frequency
ARDBP:p.Ala315Thr	ikely Pathogenic	e-06	
ARDBP:p.Ala326Thr	ikely Pathogenic	e-06	
ARDBP:p.Ser377Thr	ikely Pathogenic	.7e-05	
ARDBP:p.Ala382Thr	ikely Pathogenic	e-06	
ARDBP:p.Ile383Val	ikely Pathogenic	e-06	.572e-06
ARDBP:p.Gly386Glu	ikely Pathogenic	e-06	
ARDBP:p.Met405Val	ikely Pathogenic	.7e-05	.571e-06
BA:p.Arg535His	ikely Pathogenic	.5e-05	.528e-05
BA:p.Arg502Cys	ikely Pathogenic	.000369	.28e-05
BA:p.Val433Leu	ikely Pathogenic	e-06	.569e-06
BA:p.Asn409Ser	isk Factor/Modifier	.001789	.0018
BA:p.Arg398*	athogenic	.7e-05	
BA:p.Thr362Ile	ikely Pathogenic	.2e-05	
BA:p.Ser310Gly	ikely Pathogenic	.5e-05	.942e-05
BA:p.Pro305Leufs*31	athogenic	.7e-05	
BA:p.Ile299Thr	ikely Pathogenic	e-06	
BA:p.Arg296Gln	ikely Pathogenic	.000117	.943e-05
BA:p.Phe255Tyr	ikely Pathogenic	.2e-05	.314e-05
BA:p.Phe252Ile	ikely Pathogenic	e-06	.572e-06
BA:p.Ser235Pro	ikely Pathogenic	e-06	.596e-06
BA:p.Asn227Ser	athogenic	.2e-05	.861e-05
BA:p.Pro221Thr	ikely Pathogenic	.5e-05	
BA:p.Leu199Aspfs*62	athogenic	e-06	

GBA:p.Arg87Trp	athogenic	e-06	..972e-05
GBA:chr1:g.155240629C>T	athogenic	.7e-05	'.885e-05
SEN2:p.Lys82Ifs*44	ikely Pathogenic	e-06	
SEN2:p.Met298Thr	ikely Pathogenic	.e-05	.579e-06
SEN2:chr1:g.226895650C>A	ikely Pathogenic	e-06	.943e-05
BK1:p.Arg117*	athogenic	e-06	
BK1:chr12:g.64482022G>A	athogenic	.e-05	
BK1:p.Glu643delGlu	ikely Pathogenic	..2e-05	
SEN1:p.Ala79Val	athogenic	e-06	.286e-05
SEN1:p.Ala137Thr	ikely Pathogenic	e-06	
SEN1:p.Ile143Phe	athogenic	e-06	
SEN1:p.Ala246Glu	athogenic	e-06	
SEN1:p.Pro267Leu	athogenic	e-06	
SEN1:p.Arg269His	athogenic	..7e-05	.579e-06
SEN1:p.Val412Ile	athogenic	e-06	..972e-05
HRNB4:p.Leu60Phe	ikely Pathogenic	e-06	
GRN:p.Ser78Phefs*41	athogenic	e-06	
GRN:chr17:g.44349552G>A	athogenic	e-06	
GRN:p.Leu187Argfs*69	athogenic	e-06	
GRN:p.Cys253*fs*1	athogenic	e-06	
MAPT:p.Gly55Arg	athogenic	e-06	..314e-05
MAPT:p.Gly273Arg	athogenic	e-06	.571e-06
MAPT:chr17:g.46010403G>A	athogenic	e-06	
MAPT:chr17:g.46010418C>T	athogenic	..2e-05	
MAPT:chr17:g.46010421C>G	athogenic	..7e-05	..314e-05
MAPT:p.Gln336Arg	athogenic	e-06	
MAPT:p.Ser352Leu	athogenic	e-06	.574e-06
MAPT:p.Gly366Arg	athogenic	e-06	
MAPT:p.Gly389Arg	athogenic	e-06	
MAPT:p.Arg406Trp	athogenic	..5e-05	
MAPT:p.Asn410His	athogenic	'e-06	
MAPT:p.Thr427Met	athogenic	.e-05	..314e-05
MPO:chr17:g.58270865T>G	athogenic	.006065	.0046
MPO:p.Arg569Trp	ikely Pathogenic	.003414	.0017
MPO:p.Glu202*	athogenic	.000314	.0001
MPO:chr17:g.58280016T>C	athogenic	.0001	.0001
ABCA7:p.R1118*	ikely Pathogenic	e-06	..971e-05
ABCA7:p.Arg1561*	athogenic	..5e-05	..971e-05
ABCA7:p.Glu1679*	isk Factor/Modifier	.2e-05	.945e-05
APP:p.Val717Ile	athogenic	e-06	
APP:p.Glu693Gln	athogenic	e-06	

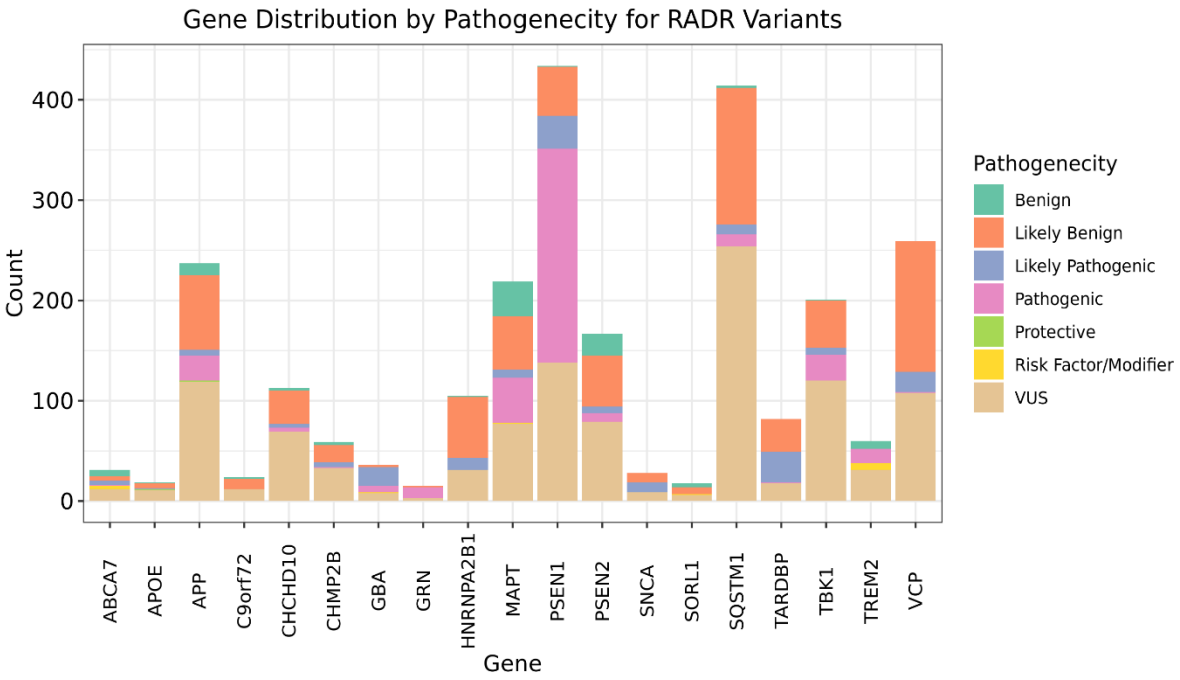
:HCHD10:p.Gln82*	'athogenic	'e-06	
:HCHD10:p.Ser59Leu	ikely Pathogenic	:e-06	
:HMP2B:p.Thr89Lysfs*2	ikely Pathogenic	'e-06	
:HMP2B:p.Ala179Profs*32	ikely Pathogenic	:e-06	
:HMP2B:p.Arg186*	ikely Pathogenic	:e-06	;.587e-06
NCA:p.Pro120Leu	ikely Pathogenic	'e-06	;.576e-06
NCA:p.Gly73Ser	ikely Pathogenic	:e-06	
IFE:chr6:g.26093233G>A	'athogenic	:e-06	..915e-05
REM2:p.Leu211Pro	:isk Factor/Modifier	..005076	..0386
REM2:chr6:g.41159790A>G	'athogenic	:e-06	
REM2:p.His157Tyr	:isk Factor/Modifier	..000192	..0019
REM2:p.Val126Gly	'athogenic	:e-06	
REM2:p.Asp87Asn	:isk Factor/Modifier	..000558	..0014
REM2:p.Thr66Met	'athogenic	..7e-05	..912e-05
REM2:p.Arg62His	:isk Factor/Modifier	..009494	..0071
REM2:p.Tyr38Cys	'athogenic	:e-06	
REM2:p.Gln33X	:isk Factor/Modifier	..2e-05	..971e-05
HNRNPA2B1:p.Tyr278Cys	ikely Pathogenic	:e-05	
'CP:p.Arg159Cys	ikely Pathogenic	:e-06	
'CP:p.Pro137Ser	ikely Pathogenic	:e-06	
'CP:p.Arg93Cys	ikely Pathogenic	'e-06	

Supplemental Table 1. BioMe Pathogenic/Likely Pathogenic/Risk Factor allele frequencies

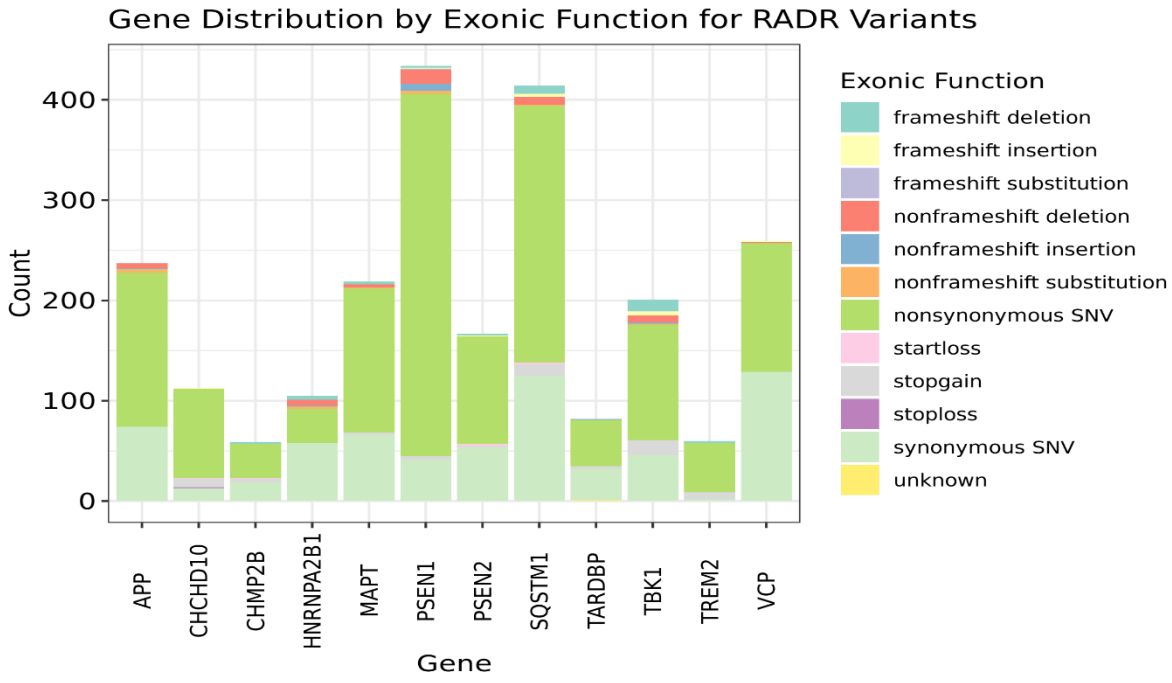
'ariant	'athogenicity	llele Frequency	Gnomad 3.12 Allele Frequency
'ARDBP:p.Gly294Val	ikely Pathogenic	..24538e-05	..314e-05
'ARDBP:p.Gly351Ser	ikely Pathogenic	..62269e-05	
;BA:p.Arg502Cys	ikely Pathogenic	;.11346e-05	'.28e-05
;BA:p.Val433Leu	ikely Pathogenic	..24538e-05	;.569e-06
;BA:p.Asn409Ser	:isk Factor/Modifier	..00436504	..0018
;BA:p.Arg398*	'athogenic	..86808e-05	
;BA:p.Ala348Gly	ikely Pathogenic	..24538e-05	;.571e-06
;BA:p.Ser310Gly	ikely Pathogenic	..24538e-05	;.942e-05
;BA:p.Phe252Ile	ikely Pathogenic	..62269e-05	;.572e-06
;BA:p.Ser235Pro	ikely Pathogenic	..62269e-05	;.596e-06
;BA:p.Asn227Ser	'athogenic	..86808e-05	..861e-05
;BA:p.Leu199Aspfs*62	'athogenic	..62269e-05	
;BA:p.Arg87Trp	'athogenic	;.49077e-05	..972e-05
;BA:p.Thr75delThr	'athogenic	..000113588	..0001
'BK1:p.Ile141Asnfs*22	'athogenic	..62274e-05	;.574e-06
'BK1:p.Ser151Cys	ikely Pathogenic	..24538e-05	..315e-05
'BK1:p.Glu643delGlu	ikely Pathogenic	..62311e-05	

'SEN1:p.Ala79Val	'athogenic	..62269e-05	..286e-05
'SEN1:p.Gly206Ala	'athogenic	..000778892	..0001
'SEN1:p.Val412Ile	'athogenic	..24538e-05	..972e-05
'PS13C:p.Arg3652*	'athogenic	..00016229	..0002
/APT:p.Ala152Thr	isk Factor/Modifier	..00133061	..0018
/APT:p.Pro397Ser	'athogenic	..24538e-05	..571e-06
/PO:p.Arg569Trp	ikely Pathogenic	..00115211	..0017
/PO:p.Glu202*	'athogenic	..49077e-05	..0001
\BCA7:p.Pro1261Leufs*112	isk Factor/Modifier	..62269e-05	
\BCA7:p.Arg1561*	'athogenic	..62269e-05	..971e-05
\BCA7:p.Glu1679*	isk Factor/Modifier	..24538e-05	..945e-05
NCA:p.Glu123Lys	ikely Pathogenic	..62269e-05	..574e-06
QSTM1:p.Gln354Valfs*37	ikely Pathogenic	..62269e-05	
HFE:p.Cys282Tyr	ikely Pathogenic	..0221822	..0373
'REM2:p.Leu211Pro	isk Factor/Modifier	..0526888	..0386
'REM2:p.His157Tyr	isk Factor/Modifier	..00379735	..0019
'REM2:p.Asp87Asn	isk Factor/Modifier	..000730211	..0014
'REM2:p.Thr66Met	'athogenic	..86808e-05	..912e-05
'REM2:p.Arg47His	isk Factor/Modifier	..00345633	..0020
'REM2:p.Arg47Cys	'athogenic	..62269e-05	..57e-06
'REM2:p.Tyr38Cys	'athogenic	..62269e-05	
'REM2:p.Gln33X	isk Factor/Modifier	..24538e-05	..971e-05
'CP:p.Pro137Ser	ikely Pathogenic	..62269e-05	

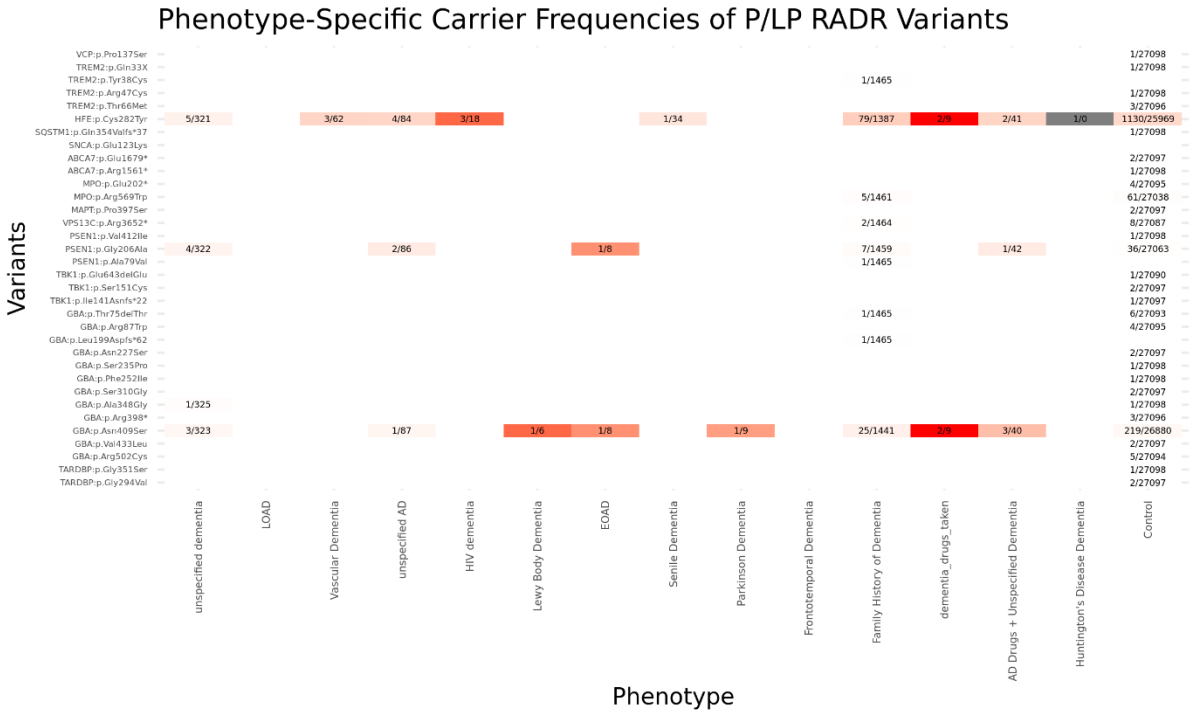
Supplemental Table 2. UKB Pathogenic/Likely Pathogenic/Risk Factor Allele Frequencies



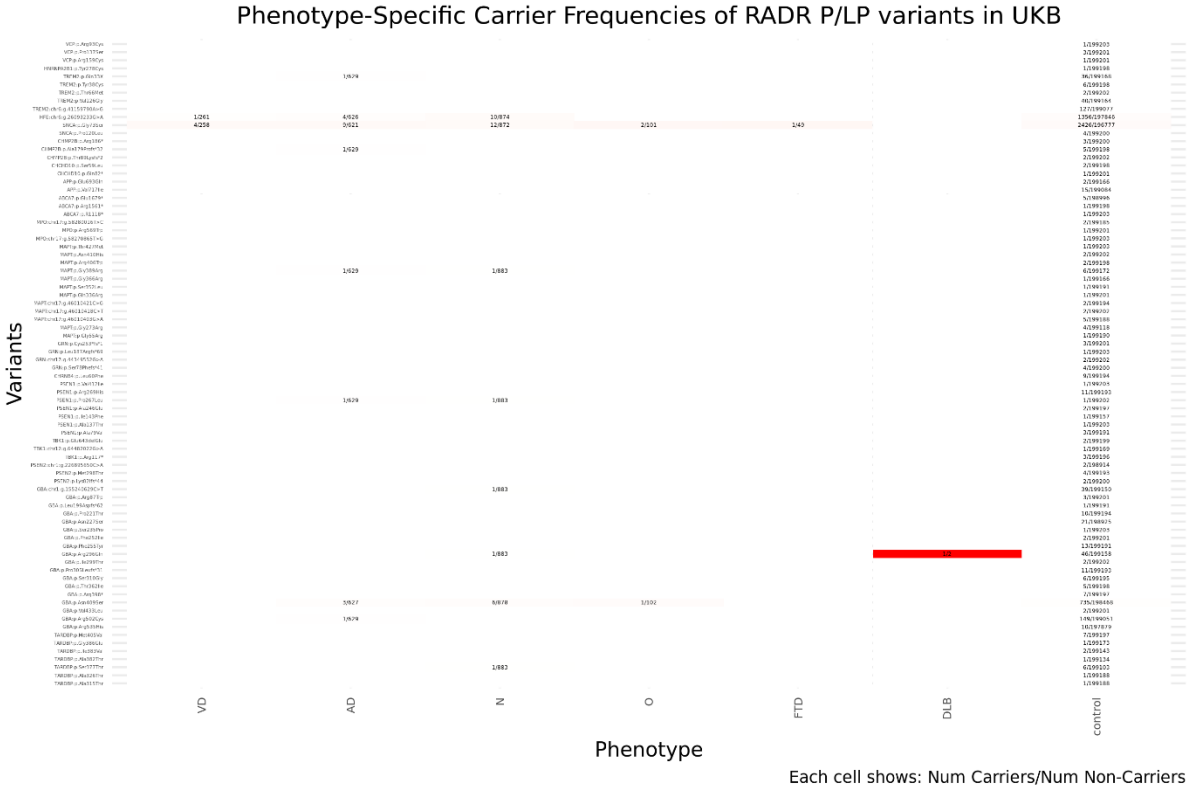
Supplemental Figure 1. Distribution of RADR Genes stratified by exonic function



Supplemental Figure 2. Distribution of RADR Genes stratified by pathogenicity.



Supplemental Figure 3. Heatmap of all phenotype-specific carrier and non carrier rates of RADR pathogenic and likely pathogenic variants found in BioMe



Supplemental Figure 4. Heatmap of all phenotype-specific carrier and non carrier rates of RADR pathogenic and likely pathogenic variants found in UKB.