

# Alzheimer's Disease Variant Portal: A Catalog of Genetic Findings for Alzheimer's Disease

Pavel P. Kuksa<sup>a,b,1</sup>, Chia-Lun Liu<sup>a,b,1</sup>, Wei Fu<sup>a,b,1</sup>, Liming Qu<sup>a,b</sup>, Yi Zhao<sup>a,b</sup>, Zivadin Katanic<sup>a,b</sup>, Kaylyn Clark<sup>a,c</sup>, Amanda B. Kuzma<sup>a,b</sup>, Pei-Chuan Ho<sup>a,b</sup>, Kai-Teh Tzeng<sup>d</sup>, Otto Valladares<sup>a,b</sup>, Shin-Yi Chou<sup>a,b,d</sup>, Adam C. Naj<sup>a,c</sup>, Gerard D. Schellenberg<sup>a,b</sup>, Li-San Wang<sup>a,b,\*,1</sup> and Yuk Yee Leung<sup>a,b,\*,1</sup>

<sup>a</sup>*Penn Neurodegeneration Genomics Center, Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia PA, USA*

<sup>b</sup>*Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia PA, USA*

<sup>c</sup>*Department of Biostatistics, Epidemiology, and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia PA, USA*

<sup>d</sup>*Department of Economics, Lehigh University, Bethlehem, PA, USA*

<sup>e</sup>*Genomics and Computational Biology Graduate Group, Perelman School of Medicine, University of Pennsylvania, Philadelphia PA, USA*

Handling Associate Editor: Anthony Griswold

Accepted 17 December 2021

Pre-press 18 January 2022

## Abstract.

**Background:** Recent Alzheimer's disease (AD) genetics findings from genome-wide association studies (GWAS) span progressively larger and more diverse populations and outcomes. Currently, there is no up-to-date resource providing harmonized and searchable information on all AD genetic associations found by GWAS, nor linking the reported genetic variants and genes with functional and genomic annotations.

**Objective:** Create an integrated/harmonized, and literature-derived collection of population-specific AD genetic associations.

**Methods:** We developed the Alzheimer's Disease Variant Portal (ADVP), an extensive collection of associations curated from >200 GWAS publications from Alzheimer's Disease Genetics Consortium and other consortia. Genetic associations were systematically extracted, harmonized, and annotated from both the genome-wide significant and suggestive loci reported in these publications. To ensure consistent representation of AD genetic findings, all the extracted genetic association information was harmonized across specifically designed publication, variant, and association categories.

---

<sup>1</sup>These authors contributed equally to this work.

\*Correspondence to: Li-San Wang, PhD, Richards Building, D101, 3700 Hamilton Walk, Philadelphia, PA 19104, USA. Tel.: +1 215 746 7015; Fax: +1 215 573 3111; E-mail: lswang@pennmedicine.upenn.edu.; and Yuk Yee Leung, PhD, Richards Building, D102, 3700 Hamilton Walk, Philadelphia, PA 19104, USA. Tel.: +1 215 573 3729; Fax: +1 215 573 3111; E-mail: yyee@pennmedicine.upenn.edu.

**Results:** ADVP V1.0 (February 2021) catalogs 6,990 associations related to disease-risk, expression quantitative traits, endophenotypes, or neuropathology. This extensive harmonization effort led to a catalog containing >900 loci, >1,800 variants, >80 cohorts, and 8 populations. Besides, ADVP provides investigators with a seamless integration of genomic and publicly available functional annotations across multiple databases per harmonized variant and gene records, thus facilitating further understanding and analyses of these genetics findings.

**Conclusion:** ADVP is a valuable resource for investigators to quickly and systematically explore high-confidence AD genetic findings and provides insights into population-specific AD genetic architecture. ADVP is continually maintained and enhanced by NIAGADS and is freely accessible at <https://advp.niagads.org>.

**Keywords:** AD GWAS literature curation, Alzheimer's disease, data curation, database, genome-wide association studies, harmonization

## INTRODUCTION

Alzheimer's disease (AD) is a devastating neurological disorder affecting millions of people worldwide and is the most common cause of dementia [1]. There are no approved drugs that can slow or treat the disease. The disease is complex and highly heritable [2]. The strongest known genetic risk factor for AD is the  $\epsilon 4$  allele of the Apolipoprotein E gene (*APOE*  $\epsilon 4$ ) [3, 4], but more than one-third of AD cases do not carry any *APOE*  $\epsilon 4$  alleles. Large-scale genome-wide association studies (GWAS) have led to the discovery of additional common genetic loci associated with the late-onset AD (LOAD) [5–9]. Yet, the identification of genetic contributors to LOAD remains a challenge as LOAD is likely caused by multiple low penetrance genetic variants [10], with the small sample sizes further complicating the identification of these causal variants.

The Alzheimer's Disease Genetics Consortium (ADGC) was founded in 2009 and funded by National Institute on Aging (NIA), to conduct large sample GWAS to identify genes associated with an increased risk of developing LOAD. ADGC co-founded IGAP (International Genomics of Alzheimer's Project) with three other AD genetics consortia: Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium, the European Alzheimer's Disease Initiative (EADI), and the Genetic and Environmental Risk in Alzheimer's Disease (GERAD) Consortium. IGAP assembled large Caucasian samples for better statistical power and was able to identify 19 genome-wide significant loci in 2013 [11], and five more loci using 30,000 samples more in total in 2019 [12].

In addition to GWAS focusing on association with disease risk, many recent genetic studies have been performed on related phenotypes including neuroimaging biomarkers [13], circulating biomarkers [14, 15], cognitive decline [16, 17], neuropathology

[18], and family history [19]. GWAS on Hispanic, African-American, Asian, and other minority populations also led to new variants not observed in Caucasians [20–22].

In order to help investigators better explore the rich and diverse literature of genetic findings, it is important to have a single resource with harmonized, unified, searchable information on identified genetic variants and genes across a variety of AD studies and populations, along with supporting functional genomic evidence.

To meet this need, we have cataloged genome-wide significant and suggestive genetic association results from all major GWAS published by ADGC (2009–2020) and other AD GWAS publications from the NHGRI/EBI GWAS Catalog [23]. Summaries extracted from each of the articles are made publicly available on the Alzheimer's Disease Variant Portal (ADVP) (<https://advp.niagads.org>). To date, ADVP provides the largest, most updated, and comprehensive collection of systematically curated, harmonized, and annotated AD-specific genetic associations. This first release contains information on 6,990 genetic associations and >900 loci curated and harmonized from >200 AD publications. ADVP will serve as an invaluable resource for the research community to explore and decipher the genetic architecture of AD and other neurodegenerative diseases.

## MATERIALS AND METHODS

An overview of the ADVP study design is shown in Fig. 1. Supplementary Note 1 provides details on ADVP architecture and implementation.

### *Data collection: Collection and curation of AD-related GWAS publications*

This ADVP V1.0 release consists of curated and harmonized genetic associations from the

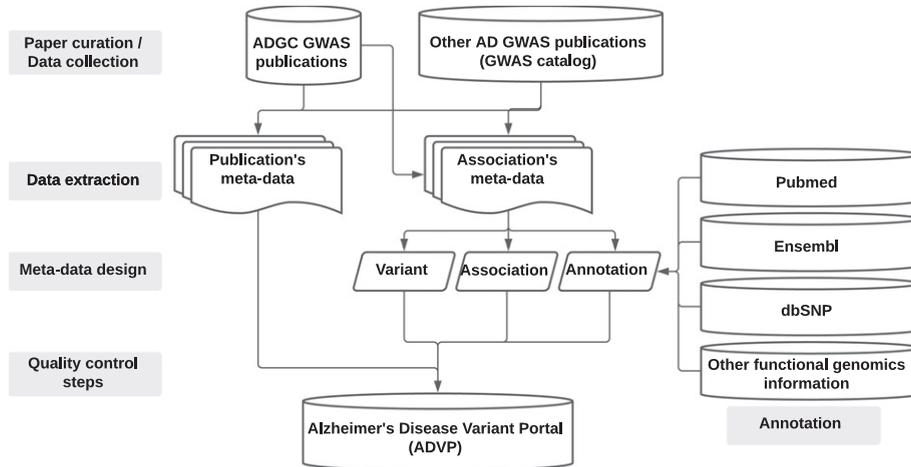


Fig. 1. ADVP study design. AD GWAS publications are first collected (Section “Data collection”), genetic variant and association data are then systematically extracted (Section “Data extraction”), harmonized (Section “Meta-data design”), annotated (Section “Annotation”), subjected to quality control steps (Section “Quality control steps”), and stored into ADVP.

genome-wide significant and suggestive loci collected from AD genetic studies conducted primarily by the ADGC. ADGC identified and compiled peer-reviewed publications which used ADGC data by searching the PubMed database for publications with the Consortium’s name as a coauthor or in the acknowledgments, or in which the ADGC’s NIA grant number (U01AG032984) is referenced or acknowledged. Use of data was confirmed through the methods or supplemental section of the publication. The ADGC AD GWAS publications (2009-January 2020, <https://www.adgenetics.org>) together with the AD GWAS studies curated using “Alzheimer’s Disease” (MeSH D000544) in GWAS catalog [23] (curation date: January 2020) and not in the ADGC set, were included in the initial set of publications for further curation. All the initial publications (total  $N = 205$ ; primary ADGC:  $N = 134$ ; additional publications identified from ADGC review papers:  $N = 20$ ; additional publications from GWAS catalog:  $N = 51$ ) were first screened to identify publications reporting GWAS findings. For each publication reporting GWAS findings, all genetic associations reported in the main text (table format) were systematically extracted. As the current ADVP release focuses on the major/main findings in each publication/GWAS study, we did not extract associations reported in the supplementary tables as they often represent supplemental findings not reaching the same level of statistical significance as the main associations. In total, we curated 125 publications which contained GWAS findings reported in the main text (<https://advp.niagads.org/publications>). Supple-

mentary Table 1 provides details on all curated AD publications in ADVP V1.0. Note the ADGC family-based analyses results will be included in the next release.

#### *Data extraction: Extraction of genetic variants and associations from publications*

We applied the following systematic data extraction and curation procedure for each publication to organize all the extracted variant and association information into a structured tabular format according to the pre-specified/designed ADVP data schema (see Section “Meta-data design” for details about the ADVP schema/columns). In each publication, we identified all the tables in the main text with reported association  $p$ -values. All the information for these associations was then saved into corresponding template publication meta-data and association meta-data documents. We then performed harmonization for few ADVP fields to ensure consistency and reduce variability when presenting extracted/reported information across publications (e.g., cerebrospinal fluid amyloid beta (1-42) levels can be reported as “CSF  $A\beta_{1-42}$ ” or as “CSF A $\beta$ ”; we map all such reported terms to the same “CSF A $\beta_{1-42}$ ” term). A second curator was used to verify if the publication records match the curated contents made by the first curator.

Lastly, the extracted information was parsed by customized scripts to validate, annotate, and store the publication, variant, and association data in the relational database [33].

### *Meta-data design: Meta-data schema for systematic curation and harmonization of genetic associations*

#### *Publication meta-data*

Meta-data for all curated publications in ADVP was extracted from PubMed (<https://pubmed.ncbi.nlm.nih.gov>) using the NCBI EDirect command-line interface (<https://www.ncbi.nlm.nih.gov/books/NBK179288>) with publication PubMed identifiers (PMID) as query keys. For each publication, we record its PMID, PubMed Central identifier (PMCID), first and last authors, journal, and year of publications. We also store the abstract, article URLs, and information on curated/source tables in the Publication meta-data (Fig. 1).

#### *Association meta-data*

ADVP association meta-data consists of 28 data fields, of which 19 are extracted directly from the paper contents. The rest of the fields are harmonized (based on extracted original information) or programmatically generated. Harmonized fields were constructed to ensure consistency and reduce variability when combining and presenting the reported/extracted information across publications, e.g., by assigning reported data values into pre-defined phenotype and population groups. Altogether, association meta-data provide 1) variant information (Section “Description of Variants”); 2) association information (Section “Description of association records”); and 3) annotation information (Section “Annotation”). For a detailed explanation of these curated and harmonized/derived data fields, see Supplementary Table 2.

#### *Description of variants*

Each genetic variant in ADVP is described using dbSNP rsID, genomic coordinates (chromosome:basepair), reference, and alternative alleles. Both the information reported in the publication (if available) and those derived from the reference databases such as dbSNP [24] and Ensembl [25] are included in the variant description. Genomic location in ADVP is currently stored using both GRCh37/hg19 and GRCh38/hg38 reference genome builds. As most associations/variants that were reported in the curated publications used GRCh37/hg19 genome build, dbSNP b155 was used to provide GRCh38/hg38 reference coordinates for

all the variants based on the reported rsID information.

#### *Description of association records*

The primary association information was systematically extracted from each source table and recorded as part of the ADVP association record. The extracted information was further recoded and categorized to ensure consistent record descriptions across publications. For each reported association, we first collected a pre-defined set of data attributes commonly reported by genetic association studies (See “Extracted” columns under Supplementary Table 2). These include variant identifier, genomic position, association *p*-value and statistics related to the effect size (regression beta coefficients and variance, odds ratios, confidence intervals), population/cohort, association model, sample size, reported effect allele and its frequency in the studied population. Collected association model information includes analysis type/model used (such as linear or logistic regression, inverse-variance weighted fixed-effect meta-analysis), any reported covariates/adjustments (such as gender, age, PCs, and *APOE*  $\epsilon$ 4 dosage), and any other details if reported by the publication.

In addition to the information directly extracted from publications, each association in ADVP is described with the nine specifically designed harmonized meta-information data fields (see Supplementary Note 2 for more details on each of these fields; field names are denoted with the double quotes below):

1. “Record type”: association record type such as SNP-based and gene-based.
2. “Population”: study population categories such as Caribbean-Hispanic, Asian, and others.
3. “Cohort”: harmonized cohort names.
4. “Sample size”: original sample size.
5. “Subset analyzed”: description of the subset of samples used to perform the association analysis.
6. “Phenotype”: the outcome variable (phenotype/trait) of the association analysis. All the curated and harmonized phenotypes are available in Supplementary Table 5.
7. “Association Type”: association type categories such as disease-risk, eQTL, endophenotype, and others.
8. “Stage”: analysis stage (discovery, validation, replication, or meta- or joint- analysis). Please

see (Supplementary Figure 1, Supplementary Note 2) for details on how this information was derived.

#### 9. "Imputation": imputation panel information.

Note, specifically, "Population", "Cohort", and "Phenotype" information are displayed in ADVP using both the original (reported) and the derived, harmonized data fields.

#### *Annotation: Genomic and functional annotation for harmonized variants and associations*

All variants and associations in ADVP were systematically annotated with genomic context (closest upstream/downstream genes), genomic element (promoter, UTR, intron, exon, intergenic, repeat), functional impact (variant most severe consequence), and cross-referenced with reference databases (dbSNP [24], GenomicsDB [26], PubMed (<https://pubmed.ncbi.nlm.nih.gov>), and Ensembl [25]).

ADVP reports the genomic context of each genetic variant via multiple data fields: 1) "Locus" – records the gene name as reported in the publication; 2) "Nearest gene" – contains the name of the gene closest to the variant and the distance to the gene (in base pairs (bps)) in upstream (+) or downstream (–) orientation. The nearest genes are identified using GENCODE v34 [27] protein-coding gene annotations. For each genetic variant co-localized with one or more genes, both EnsemblID [25] and HGNC [28] symbols for the gene(s) are reported. For each ADVP variant, the co-localized genomic element is reported based on the genomic partition information [29, 30] such as exonic, intronic, promoter, intergenic, and other genomic elements.

#### *Quality control steps: Variant and association data verification*

Quality control for the variant and association information in ADVP is carried out at multiple levels:

1. We ensured records are not double-counted/reported across studies. Each association record in ADVP is uniquely identified by a combination of reported locus/variant/interaction name, cohort/analyzed subset, association model used, phenotype, and association  $p$ -value and effect size.
2. We cross-checked recorded positional information (chromosome:basepair), rsID, and allele

information against reference databases including dbSNP [24], NIAGADS Alzheimer's Genomics database [26], and 1000 Genome data [31] to ensure correctness of the variant information.

3. We identified and removed any supporting association records (e.g., records that were directly retrieved from the other resources such as GTEX [32] and used for comparison or to support primary findings).

#### *Population-based analysis of AD associations*

To understand the genetic architecture of AD associated loci across populations, we compared their reported effect sizes and frequencies for each locus. Note that the effect size values used were those reported in the publications, and we did not attempt to resolve inconsistent results within the same population. This is due to the following two challenges. First, overlapping cohorts were used across different GWAS studies. Second, there is a lack of available genome-wide summary statistics. Therefore, in this analysis we only investigated the effect directions for the top (most significant) association at each AD-associated loci per population.

To do so, we used all association records from case-control AD GWAS studies with complete information on reported allele, effect size (odds ratio), and allele frequency. We then analyzed the four major populations with the most association records (African American, Asian, Caribbean Hispanic, and Caucasian/Non-Hispanic White). Any association records with  $p > 0.01$  were excluded from the analysis. Then for each AD-associated locus (identified using the nearest gene), we selected the most significant (smallest)  $p$ -value as the representative association record for that locus in each population.

To investigate the association results across populations, we used the odds ratios and frequencies of the minor alleles since most of the analyzed publications reported the effect and frequency for minor alleles and did not always report information for both alleles. Note that ADVP itself does not use minor allele-based normalization, rather it stores reported alleles and their respective effect sizes and frequencies. We further categorized the AD-associated loci as population-specific or shared if they were associated with AD in more than one population under study. To investigate population specificity of AD loci, we then compared AD-associated loci (both shared and population-specific) across populations

based on their normalized effect sizes (odds ratios for minor alleles) and the minor allele frequency.

### Functional analysis of AD associations

To validate and assess the quality of association data stored in ADVP, we performed functional analysis of the ADVP variants. We focused this assessment on the non-eQTL AD/ADRD associations (i.e., variants with unknown functions). Therefore, for this analysis, we excluded all curated ADVP eQTL records reported in publications.

First, we evaluated the significance of overlaps between ADVP variants and active enhancer elements across tissues and cell types. We used Roadmap Epigenomics [33] (ChromHMM-determined [34]) and FANTOM5 [35] enhancer sets across 35 tissue/cell type categories. We then ranked individual tissue/cell types based on the degree and significance (odds ratio, Fisher's exact test) of enhancer overlaps in each of the analyzed tissues/cell types. Second, to further investigate potential functional roles of ADVP variants, we overlapped ADVP variants with the AMP-AD eQTL records for the three brain regions.

## RESULTS

In order to focus on association findings with the highest confidence, we decided to concentrate on

large-scale association studies at the genomic level, with the majority of studies included in ADVP (65%) reporting associations reaching genome-wide significance, a gold standard for human genetic discoveries. Furthermore, ADVP collected extensive meta-data, including consortiums and cohorts, which were not available in both the GWAS catalog [23] and AlzGene [36], and are important for relating the results reported across publications. Finally, ADVP provides convenient links for investigators to explore biological significance of the reported variants (e.g., their genomic context, available functional genomic data, or other known associations, if any) via genomic and functional annotations (Fig. 1).

In terms of the number of curated AD-related associations and publications, ADVP provides more comprehensive AD-related association information than the NHGRI-EBI GWAS catalog [23] and AlzGene [36] resource (see Table 1 for main improvements and comparisons). ADVP significantly expands coverage of AD publications by 1) providing information on 125 GWAS publications (compared to, e.g., 69 in GWAS catalog, or 41 in AlzGene), 2) includes association information for both genome-wide significant and suggestive loci and variants, 3) contains 4 times more association records, and 4) 60% more of distinct AD-related variants compared to these resources (Table 1).

Table 1  
Comparison of the contents and main features between ADVP and other genetic databases

Contents/Features	AlzGene [36]	GWAS Catalog [23]	ADVP (this work)
Number of curated AD association records	Not reported	1,532	<b>6,990</b>
Number of curated AD genetics publications	41	69	<b>125</b>
Number of reported variants	1,055	1,155	<b>1,786</b>
Number of reported genes	264	508	<b>984</b>
Variant and gene association records	No (variant only)	No (variant only)	Yes (variant, gene, SNP-SNP, and gene-gene interactions)
Genome-wide significant and suggestive associations	No (genome-wide significant only, $p < 5 * 10^{-8}$ )	No (genome-wide significant only, $p < 5 * 10^{-8}$ )	Yes (both genome-wide significant and suggestive associations)
Genomic and functional annotations for AD-related variants and genes	No	No	Yes (genomic context, affected genomic elements, variant effect, and other annotations)
Curated and harmonized population, cohort, phenotype, association type information	No	No	Yes
Imputation/imputation panel information	No	No	Yes
Last update date	2011	2021	2021

\*Highest numbers of records are highlighted in bold.

Following the ADVP curation criteria (see “Data collection”, Fig. 1), we first identified and screened 205 AD-related publications from 2009 to January 2020. Out of these, we identified 125 publications with genetic associations reported in the main text tables ( $N=225$  tables). Genetic variant and association data were then systematically extracted (Section “Data extraction”), harmonized (“Meta-data design”), annotated (“Annotation”), subjected to quality control steps (“Quality control steps”), and stored in ADVP (Fig. 1).

#### *ADVP data summary*

The ADVP V1.0 release contains high-quality genome-wide and suggestive AD-related genetic associations extracted from GWAS publications. This includes 6,990 genetic associations for variants, genes, and SNP interactions. Figure 2 shows the distribution of ADVP genetic associations by harmonized meta-information data fields: a) Nine harmonized phenotypes; b) Six harmonized analyses type; c) Population, and d) Cohorts/Consortiums.

All ADVP association records are described by our uniquely designed categories:

1. As shown in Fig. 2A, ADVP records are associated with nine different phenotype categories, with roughly half of them related to AD diagnosis. 15% of the records are related to fluid biomarkers, 7% with imaging, and 6% with cognitive measures.
2. With respect to analysis type categories, ADVP includes 3,199 (45.8%) association records reported in disease-risk analyses, of which 1,342 and 934 associations are reported by meta- and joint-analyses, respectively. 1,887 (26.9%) of the records are related to AD endophenotype and 924 (13.2%) eQTL AD associations (Fig. 2B).
3. ADVP is the first to collect AD genetic associations at SNP level (6,437, 92.1%), gene level (320, 4.5%), as well as SNP and gene interactions (233, 3.3%).
4. ADVP records present analyses results from seven populations as well as those from multi-ethnic analyses. ~88% of the records are for Caucasian (Fig. 2C). Others include African American, Arab, Asian, Caribbean Hispanic, Hispanic, and Non-Hispanic Caucasian.
5. ADVP records span analyses results published by ADGC using over 80 cohorts

(Fig. 2D), including ADGC, Alzheimer's Disease Neuroimaging Initiative (ADNI), Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium, European Association of Development Research and Training Institutes (EADI), European Network for Genetic and Genomic Epidemiology (ENGAGE) Consortium, Enhancing NeuroImaging Genetics through Meta-Analysis (ENIGMA) Consortium, The International Genomics of Alzheimer's Project (IGAP), and others. See Supplementary Table 3 for details on cohorts included in ADVP.

Furthermore, ADVP provides annotation information for each genetic association (Section “Annotation”). In summary, all the genetic association records in ADVP were represented by >1,800 unique variants (based on genomic position) and >900 genomic loci. ADVP associations are mostly located in non-coding regions including intronic (52.9%), intergenic (15.2%), and promoter (5.9%) (Fig. 3A). ADVP records are also cross-referenced to NIAGADS Alzheimer's Genomics database [26]. The functional impact of genetic variants was determined by the ADSP functional annotation pipeline [26, 37, 38] with customized ranking to generate the most damaging consequence for each variant [38]. Figure 3B shows the functional impact of the genetic variants in ADVP as explained by the VEP (Variant Effect Predictor).

#### *ADVP features: search, browse, and visualize*

ADVP aims to provide a simple and unified resource to the scientific community, allowing investigators to search and browse AD genetic association information more easily. This is first done by displaying association records using a pre-selected set of most important data fields (Section “Description of association records”). Investigators can further select additional data fields via the column selector (Fig. 4A). All records are annotated with genomic and functional information, allowing investigators to explore various kinds of biological annotations (e.g., dbSNP [24], GenomicsDB [26], Ensembl [25], UCSC Genome browser [39], gnomAD [40], and functional consequences information [38, 41] for variants).

The ADVP search interface was designed based on focus group use cases. ADVP provides several ways to search for genetic association records:

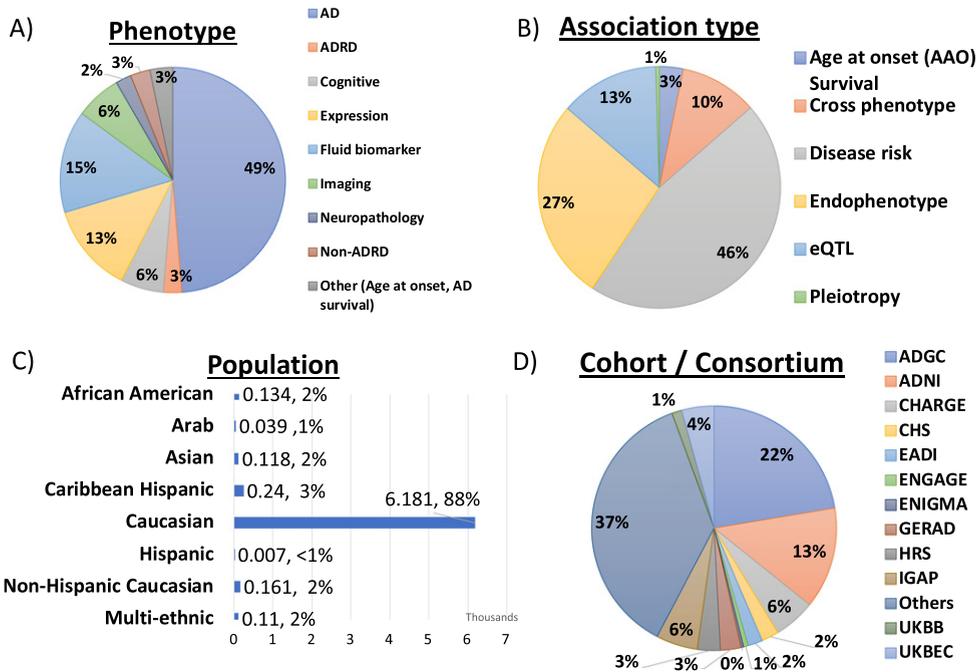


Fig. 2. Summary of genetic association records in ADVP (N=6990) by A) Phenotype, B) Association type, C) Population, and D) Cohort/Consortium.

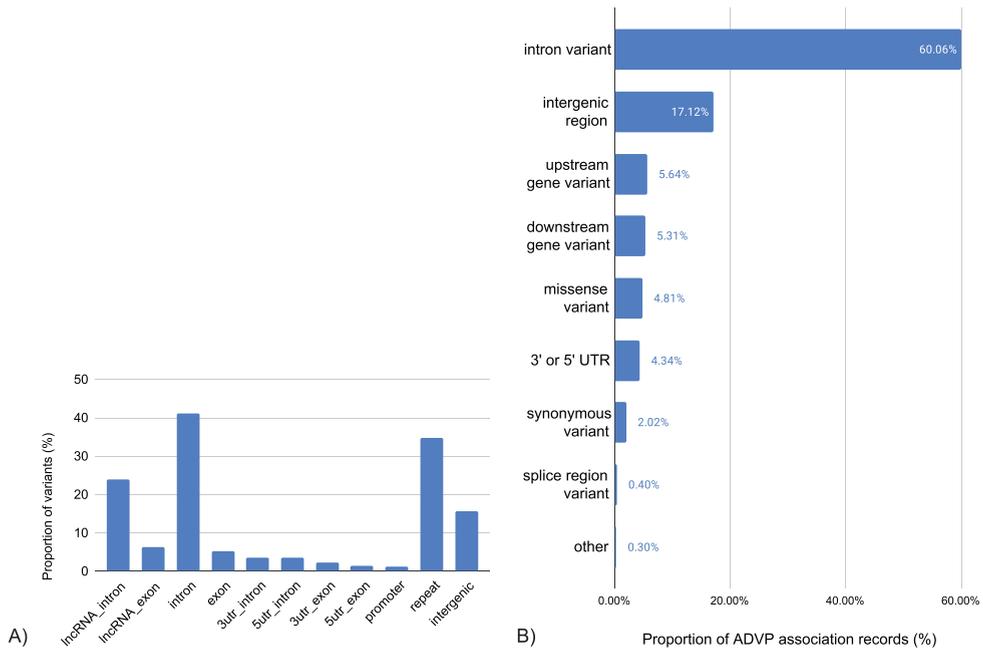


Fig. 3. Summary of ADVP association records by genome annotation and most-significant functional consequence. A) Genomic localization of ADVP variants within mRNA, lncRNA, and repeat elements. Shown are proportion (%) of variants in each genomic element category; B) Most-significant predicted variant impact. Impact for variants is determined using ADSP functional annotation pipeline [38, 41]. The consequence for a variant is predicted and ranked based on multiple criteria including genomic location of a variant, genes, transcripts and protein sequences, biological type of transcript, transcript support level and other factors.

A)

Showing 1 - 12 of 12 records

Pubmed ID	Record type	SNP	Coordinates	Locus	Population	Cohort	Sample size	Subset Analyzed	Phenotype
30820047	SNP-based	rs75932628	6:41129252	TREM2	Non-Hispanic Caucasian	ADGC, CHARGE, EADI, GERAD	63,926	All	AD
30820047	SNP-based	rs75932628	6:41129252	TREM2	Non-Hispanic Caucasian	ADGC, CHARGE, EADI, GERAD	82,771	All	AD
23150908	SNP-based	rs75932628	6:41129252	TREM2	Caucasian	Cohort from Europe	113,600	All	AD
23150908	SNP-based	rs75932628	6:41129252	TREM2	Caucasian	Cohort from Europe	12,438	All	AD
23150908	SNP-based	rs75932628	6:41129252	TREM2	Caucasian	Cohort from Europe	4,786	All	AD
28714976	SNP-based	rs143332484	6:41129207	TREM2	Caucasian	CHARGE, EADI, GERAD	3,968	All	AD
28714976	SNP-based	rs75932628	6:41129252	TREM2	Caucasian	ADGC	14,884	All	AD
28714976	SNP-based	rs75932628	6:41129252	TREM2	Caucasian	ADGC, CHARGE, EADI, GERAD	30,018	All	AD
28714976	SNP-based	rs143332484	6:41129207	TREM2	Caucasian	ADGC, CHARGE, EADI, GERAD	33,786	All	AD
28714976	SNP-based	rs75932628	6:41129252	TREM2	Caucasian	CHARGE, EADI, GERAD	35,831	All	AD
28714976	SNP-based	rs143332484	6:41129207	TREM2	Caucasian	ADGC, CHARGE, EADI, GERAD	53,042	All	AD
28714976	SNP-based	rs75932628	6:41129252	TREM2	Caucasian	ADGC, CHARGE, EADI, GERAD	80,733	All	AD

Page 1 of 1 | Show 20 records

B)

## Variants

Sort by frequency

Showing 1 - 10 of 1826 records

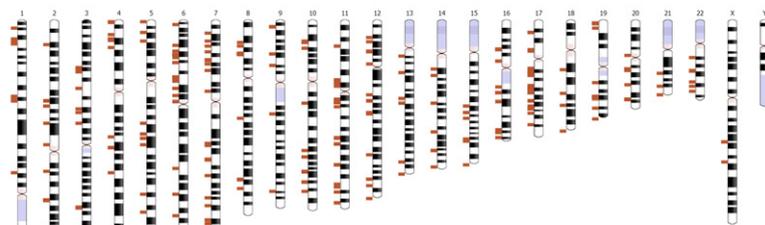
Variant	Reported gene(s)	Location	View in Genomics	Association records	Papers
rs3865444	CD33	chr19:51727962	<a href="#">View</a>	<a href="#">39</a>	<a href="#">16</a>
rs11136000	CLU; CUGBP2; PICALM	chr8:27464519	<a href="#">View</a>	<a href="#">60</a>	<a href="#">15</a>
rs2075650	APOC1, APOC4, APOE, PV	chr19:45395619	<a href="#">View</a>	<a href="#">43</a>	<a href="#">15</a>
rs11218343	PICALM; SORL1	chr11:121435587	<a href="#">View</a>	<a href="#">27</a>	<a href="#">13</a>
rs3851179	CUGBP2; PICALM	chr11:85868640	<a href="#">View</a>	<a href="#">50</a>	<a href="#">13</a>
rs6656401	CR1, CR2; CR1; CR1L; CR2	chr1:207692049	<a href="#">View</a>	<a href="#">76</a>	<a href="#">12</a>
rs3818361	CR1; CUGBP2	chr1:207784968	<a href="#">View</a>	<a href="#">43</a>	<a href="#">11</a>
rs157580	PVRL2, TOMM40; TOMM4	chr19:45395266	<a href="#">View</a>	<a href="#">15</a>	<a href="#">10</a>
rs1476679	PILRA, PILRB; GATS; MSX	chr7:100004446	<a href="#">View</a>	<a href="#">131</a>	<a href="#">9</a>
rs405509	APOE, TOMM40; APOE	chr19:45408836	<a href="#">View</a>	<a href="#">28</a>	<a href="#">9</a>

Page 1 of 183 | Show 10 records

C)

## Chromosome ideogram

To browse variants, click on a particular chromosome for enlarged view of the chromosome and associated genomic loci and/or click on genomic locus marker (red rectangle) to display variant table for a particular genomic locus. When in chromosome view, click on the chromosome again, to go back to the ideogram.



Legend  
■ ADVP genome-wide significant variants (p ≤ 5e-8)

Fig. 4. (Continued)

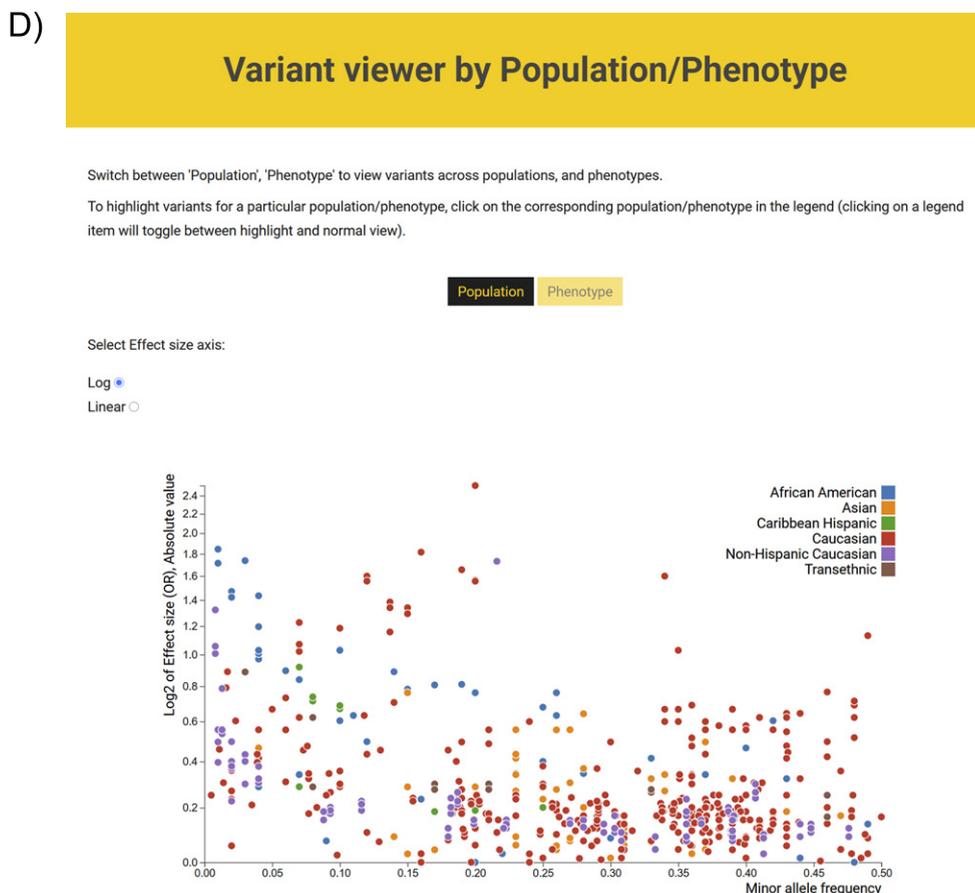


Fig. 4. ADVP interface. A) Association records table. The displayed information can be customized via column/field selector and filtered using provided text and data filters; B) Top variants curated in ADVP. Variants are displayed according to the number of reporting publications by default; association records for variants and variant-related publications can be quickly accessed; C) Interactive chromosome ideogram-based view of association data; D) Interactive variant viewer by population and phenotype. Variants are arranged by their effect size (odds ratio; Y-axis) and allele frequency (X-axis) and color-coded by population and phenotype.

1. By publication – investigators can quickly identify and retrieve all association records curated by ADVP for a particular study using PMID, PMCID, first or last author names, year of publication or article title (<https://advp.niagads.org/publications>).
  2. By variant or gene of interest – investigators can search for the variant (<https://advp.niagads.org/variants>) or gene (<https://advp.niagads.org/genes>) of interest and browse all the associated ADVP records. Additionally, investigators can easily discover top variants or a gene with most association records or most publications via the summary counts for association records and papers (Fig. 4B, Supplementary Figure 2A).
  3. By region of interest – investigators can search and retrieve all genetic associations within the genomic regions of interest (<https://advp.niagads.org/search>).
  4. By integrative genome-wide plots – investigators can navigate the landscape of AD genetics associations using the interactive chromosome ideogram ([https://advp.niagads.org/chromosome\\_ideogram](https://advp.niagads.org/chromosome_ideogram), Fig. 4C) or interactive population/phenotype variant viewer (<https://advp.niagads.org/plot>, Fig. 4D).
- Additionally, for each variant or gene, users can view the biological annotations and their functional evidence (e.g., dbSNP [24], gnomAD [40], UCSC Genome Browser [39], Ensembl [25], GenomicsDB [26]) (Supplementary Figure 2B).

### ADVP use cases

ADVP is designed with various use cases in mind. First, ADVP can serve as a point of entry for investigators to explore the AD genetics literature. They can browse through variant and gene records, identify top associated loci for particular populations and phenotypes, or inspect top GWAS associations in the gene or genomic region of interest. Second, investigators can use ADVP to check their association analysis findings. They can further restrict the comparison by focusing on results from a specific population, cohort, or by comparing the strength of associations via *p*-values or phenotypes. Lastly, investigators can use ADVP to check if their findings have functional support from eQTL or other biological annotations (Supplementary Figure 2B).

### Genetic architecture of AD-associated loci across populations

To show the diversity and breadth of ADVP data, we performed population-based analysis of AD associations in ADVP (Section “Population-based analysis of AD associations”). Across four major populations (African American, Asian, Caribbean Hispanic, Caucasian/Non-Hispanic white) with the most association records, 91 loci (Section “Population-based analysis of AD associations”) were identified in any of these 4 populations, whereas 10 of them were found in two or more populations, including *BIN1*, *CD33*, *PICALM*, *SORL1*, and *ABCA7*. The majority of AD loci ( $81/91 = 89\%$ ) were population-specific, i.e., found in only one population. This could partially be explained by the underlying genetic differences across populations but could also be contributed by variability in GWAS sample sizes across studies/populations, which could lead to the observed differences in association strength and heterogeneity of loci identified in each population. Across the four populations, the corresponding top SNP (Section “Population-based analysis of AD associations”), effect size and allele frequency on all these 91 loci are available in Supplementary Table 4.

We next explored the effect sizes and allele frequencies of all the AD-associated loci that were found in any of the populations (Fig. 5A; only non-*APOE* loci from 4 populations are shown), or those that were shared (i.e., found in two or more populations (Fig. 5B)). ADVP data (Fig. 5) captures the diversity and population-specificity of the AD-associated loci.

On one hand, the shared loci (Fig. 5B) tend to preserve their risk (e.g., *ABCA7* [42]) or protective (e.g., *PICALM* [43]) roles. On the other hand, common loci (Fig. 5A) tend to exert smaller effects compared to population-specific loci, yet the effects of the common loci vary across populations (Fig. 5B) (i.e., are population-specific).

### Functional analysis of ADVP variants

We next investigated the functional roles of ADVP variants. To do so, we analyzed all ADVP variants that are non-eQTL AD/ADRD associations (i.e., variants with unknown functional roles) in the Caucasian or non-Hispanic White populations. A total of 1,675 ADVP variants met these criteria and were analyzed.

First, to explore potential functional roles for the variants and potential causal genes for these variants, we asked if these variants regulate any genes in any of AMP-AD [44] eQTL datasets [45] obtained from three different brain regions (dorsolateral prefrontal cortex, cerebellum and temporal cortex) (see Materials and Methods and publication [45] for details of how this eQTL data was harmonized and processed).

31% of the analyzed ADVP variants were identified as significant eQTLs in at least one of the three AMP-AD eQTL datasets ( $FDR < 0.01$ ), and 32% among these were eQTLs in all three brain regions. Altogether, these variants targeted 130 genes (including HLA region), of which 31 were also the nearest genes reported in ADVP. These target genes (e.g., *ACE*, *PVR*) were enriched in cell junction organization [46], and acetyltransferase activity (cholinergic system) [47], which recapitulated some of the molecular and cellular mechanisms underlying the pathogenesis of AD.

Second, we analyzed and ranked relevant tissue categories implicated by ADVP variants. To do so, we quantified the significance and enrichment of tissue-specific enhancer overlaps with ADVP variants (Section “Functional analysis of AD associations”). The ADVP variants were enriched in regulatory elements profiled by FANTOM5 and Roadmap Epigenomics (ChromHMM). Top tissues enriched in overlaps between ADVP variants and Roadmap enhancers included blood ( $OR = 3.7$ ), digestive ( $OR = 3.5$ ), brain ( $OR = 2.8$ ), and skeletal ( $OR = 2.6$ ) tissue categories that are relevant to AD. The overlaps between ADVP variants and enhancers were shown to be enriched in tissues known to be related to AD in various ways [48–54]. For instance, the implicated immunity-related blood category is in



line with recent work highlighting the role of neuroinflammation in AD pathogenesis and etiology [48, 54]. Digestive is related to the gut microbiome, which can be linked to AD behavior in mice [55]. The implicated skeletal category has also been previously linked with brain atrophy in AD [50].

In summary, analyses of genetic associations in ADVP revealed potential functional roles of the AD variants in relevant tissue/cell type context and recapitulated some of the known regulatory mechanisms underlying AD pathogenesis.

## DISCUSSION

Here, we present ADVP, a portal to search, browse, and visualize the largest collection of systematically curated, harmonized, and annotated AD-specific genetic variants and associations (~7,000 genetic associations in the current V1.0 release, February 2021). Among the main distinctive features of ADVP is the uniqueness of reporting harmonized AD variant and association information, integration with the genomic annotation, and functional information as well as extensive cohort/consortium level information.

ADVP uniquely includes associations for individual SNPs, genes, and SNP-SNP and gene-gene interactions and contains curated phenotypes that are not limited to disease risk, but also includes endophenotypes, fluid biomarkers, imaging, neuropathology, and other phenotypes. Moreover, ADVP curates and records AD and ADRD eQTL association findings (Fig. 2B).

In addition to the  $p$ -values and effect sizes reported for association records, ADVP puts particular emphasis on harmonizing meta-data curated from the publications. Both the curated and derived columns are stored in the database. These include phenotype, association type, gene symbols, study information (population, cohort, sample size, subset analyzed), and details of analyses (analyses type, imputation) (Fig. 4A). All these columns enable the investigators to interpret, compare, and view these records at different levels: phenotype (Fig. 2A), population (Fig. 2C), cohort (Fig. 2D), to name a few.

All ADVP records are annotated with the genomic context (upstream/downstream genes, and their distances) and their co-localized genomic element (Fig. 3). They are also annotated with genomic annotation and functional genomic information. The structured design of ADVP association data allows

systematic integration with other genetic, genomics, and molecular databases.

Population-based and functional analyses of AD associations revealed the genetic architecture of AD-associated loci and points to tissue-specific regulatory mechanisms for AD. However, variability in GWAS sample sizes may contribute to the observed differences in associations and loci across different populations. On the other hand, ADVP will be able to capture more associations as GWAS sample sizes continue to increase. Broadening the ADVP coverage of population groups, as well as expanding functional data types and coverage will provide further insights in complex genetic architecture and biology underlying AD.

Lastly, we made substantial efforts to ensure high-quality of ADVP data contents. First, quality control at multiple levels is performed (Fig. 1, Section "Quality control steps") to ensure the uniqueness of included genetic associations (no double counting / re-reporting of associations). Besides, variant information in ADVP has been cross-checked against other reference databases such as dbSNP.

ADVP will continuously be updated with regular versioned releases. Supported by NIAGADS [56], the Alzheimer's human genetics data repository designated by National Institute on Aging, ADVP will add new publications on AD-related GWAS and corresponding associations in an ongoing manner.

In the future, ADVP data collections will consist of a broader range of genetic results:

1. AD whole-genome/whole-exome sequencing analyses
2. AD xQTL associations, where x = protein, methylation, epigenetics marks, or other molecular traits
3. Other genetic variant types, such as insertions/deletions (indel), copy number variations (CNV), or structural variations (SV) as they become available
4. AD related dementias (ADRD) and neurodegenerative disorders

Future ADVP functionality will include further collection and addition of functional genomic evidence supporting genetic associations.

To conclude, ADVP contains the largest collection of systematically curated, harmonized, and annotated literature-derived variants for AD to the best of our knowledge. The extensive and unique features in ADVP allow investigators to easily access, inter-

pret, compare, and visualize the vast collection of AD genetics findings.

## ACKNOWLEDGMENTS

The authors thank Brian Kunkle, Elizabeth Blue, Edoardo Marcora, Laura Cantwell, Lauren Kleidermacher, and Mitchell Tang for their contributions at various stages of this work. This work was supported by the National Institute on Aging [U24-AG041689, U54-AG052427, U01-AG032984, U01AG058654]; Biomarkers Across Neurodegenerative Diseases (BAND 3) (award number 18062), co-funded by Michael J Fox Foundation, Alzheimer's Association, Alzheimer's Research UK and the Weston Brain institute.

Authors' disclosures available online (<https://www.j-alz.com/manuscript-disclosures/21-5055r2>).

## DATA AND CODE AVAILABILITY

All AD variant and association information is available through the ADVP website (<https://advp.niagads.org>). The code for processing and analysis of variant and association data is available at <https://bitbucket.org/wanglab-upenn/ADVP>.

## SUPPLEMENTARY MATERIAL

The supplementary material is available in the electronic version of this article: <https://dx.doi.org/10.3233/JAD-215055>.

## REFERENCES

- [1] Alzheimer's Association (2019) 2019 Alzheimer's disease facts and figures. *Alzheimers Dement* **15**, 321-387.
- [2] Gatz M, Pedersen NL, Berg S, Johansson B, Johansson K, Mortimer JA, Posner SF, Viitanen M, Winblad B, Ahlborn A (1997) Heritability for Alzheimer's disease: The study of dementia in Swedish twins. *J Gerontol A Biol Sci Med Sci* **52**, M117-M125.
- [3] Corder EH, Saunders AM, Strittmatter WJ, Schmechel DE, Gaskell PC, Small GW, Roses AD, Haines JL, Pericak-Vance MA (1993) Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science* **261**, 921-923.
- [4] Genin E, Hannequin D, Wallon D, Sleegers K, Hiltunen M, Combarros O, Bullido MJ, Engelborghs S, De Deyn P, Berr C, Pasquier F, Dubois B, Tognoni G, Fiévet N, Brouwers N, Bettens K, Arosio B, Coto E, Del Zompo M, Mateo I, Epelbaum J, Frank-Garcia A, Helisalimi S, Porcellini E, Pilotto A, Forti P, Ferri R, Scarpini E, Siciliano G, Solfrizzi V, Sorbi S, Spalletta G, Valdivieso F, Vepsäläinen S, Alvarez V, Bosco P, Mancuso M, Panza F, Nacmias B, Bossù P, Hanon O, Piccardi P, Annoni G, Seripa D, Galimberti D, Licastro F, Soininen H, Dartigues J-F, Kamboh MI, Van Broeckhoven C, Lambert JC, Amouyel P, Campion D (2011) APOE and Alzheimer disease: A major gene with semi-dominant inheritance. *Mol Psychiatry* **16**, 903-907.
- [5] Lambert J-C, Heath S, Even G, Campion D, Sleegers K, Hiltunen M, Combarros O, Zelenika D, Bullido MJ, Tavernier B, Letenneur L, Bettens K, Berr C, Pasquier F, Fiévet N, Barberger-Gateau P, Engelborghs S, De Deyn P, Mateo I, Franck A, Helisalimi S, Porcellini E, Hanon O, European Alzheimer's Disease Initiative Investigators, de Pancorbo MM, Lendon C, Dufouil C, Jaillard C, Leveillard T, Alvarez V, Bosco P, Mancuso M, Panza F, Nacmias B, Bossù P, Piccardi P, Annoni G, Seripa D, Galimberti D, Hannequin D, Licastro F, Soininen H, Ritchie K, Blanché H, Dartigues J-F, Tzourio C, Gut I, Van Broeckhoven C, Alperovitch A, Lathrop M, Amouyel P (2009) Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. *Nat Genet* **41**, 1094-1099.
- [6] Harold D, Abraham R, Hollingworth P, Sims R, Gerrish A, Hamshe ML, Pahwa JS, Moskva V, Dowzell K, Williams A, Jones N, Thomas C, Stretton A, Morgan AR, Lovestone S, Powell J, Proitsi P, Lupton MK, Brayne C, Rubinsztein DC, Gill M, Lawlor B, Lynch A, Morgan K, Brown KS, Passmore PA, Craig D, McGuinness B, Todd S, Holmes C, Mann D, Smith AD, Love S, Kehoe PG, Hardy J, Mead S, Fox N, Rossor M, Collinge J, Maier W, Jessen F, Schürmann B, Heun R, van den Bussche H, Heuser I, Kornhuber J, Wiltfang J, Dichgans M, Frölich L, Hampel H, Hüll M, Rujescu D, Goate AM, Kauwe JSK, Cruchaga C, Nowotny P, Morris JC, Mayo K, Sleegers K, Bettens K, Engelborghs S, De Deyn PP, Van Broeckhoven C, Livingston G, Bass NJ, Gurling H, McQuillin A, Gwilliam R, Deloukas P, Al-Chalabi A, Shaw CE, Tsolaki M, Singleton AB, Guerreiro R, Mühleisen TW, Nöthen MM, Moebus S, Jöckel K-H, Klopp N, Wichmann H-E, Carrasquillo MM, Pankratz VS, Younkin SG, Holmans PA, O'Donovan M, Owen MJ, Williams J (2009) Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nat Genet* **41**, 1088-1093.
- [7] Seshadri S, Fitzpatrick AL, Ikram MA, DeStefano AL, Gudnason V, Boada M, Bis JC, Smith A V, Carassquillo MM, Lambert JC, Harold D, Schrijvers EM, Ramirez-Lorca R, Debette S, Longstreth WT, Janssens AC, Pankratz VS, Dartigues JF, Hollingworth P, Aspelund T, Hernandez I, Beiser A, Kuller LH, Koudstaal PJ, Dickson DW, Tzourio C, Abraham R, Antunez C, Du Y, Rotter JJ, Aulchenko YS, Harris TB, Petersen RC, Berr C, Owen MJ, Lopez-Arrieta J, Varadarajan BN, Becker JT, Rivadeneira F, Nalls MA, Graff-Radford NR, Campion D, Auerbach S, Rice K, Hofman A, Jonsson P V, Schmidt H, Lathrop M, Mosley TH, Au R, Psaty BM, Uitterlinden AG, Farrer LA, Lumley T, Ruiz A, Williams J, Amouyel P, Younkin SG, Wolf PA, Launer LJ, Lopez OL, van Duijn CM, Breteler MM, Olson J, Kronmal R, Robbins J, Fried LP, Burke G, Kuller LH, Tracy R, Gottdiener J, Prineas R, Becker JT, Enright P, Klein R, O'Leary DH, Furberg C (2010) Genome-wide analysis of genetic loci associated with Alzheimer disease. *JAMA* **303**, 1832-1840.
- [8] Hollingworth P, Harold D, Sims R, Gerrish A, Lambert J-C, Carrasquillo MM, Abraham R, Hamshe ML, Pahwa JS, Moskva V, et al. (2011) Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. *Nat Genet* **43**, 429-435.
- [9] Naj AC, Jun G, Beecham GW, Wang LS, Vardarajan BN, Buros J, Gallins PJ, Buxbaum JD, Jarvik GP, Crane

- PK, et al. (2011) Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. *Nat Genet* **43**, 436-441.
- [10] Naj AC, Schellenberg GD (2017) Genomic variants, genes, and pathways of Alzheimer's disease: An overview. *Am J Med Genet B Neuropsychiatr Genet* **174**, 5-26.
- [11] Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, DeStafano AL, Bis JC, Beecham GW, Grenier-Boley B, et al. (2013) Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet* **45**, 1452-1458.
- [12] Kunkle BW, Grenier-Boley B, Sims R, Bis JC, Damotte V, Naj AC, Boland A, Vronskaya M, van der Lee SJ, Amlie-Wolf A, et al. (2019) Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates A $\beta$ , tau, immunity and lipid processing. *Nat Genet* **51**, 414-430.
- [13] Biffi A, Anderson CD, Desikan RS, Sabuncu M, Cortellini L, Schmansky N, Salat D, Rosand J (2010) Genetic variation and neuroimaging measures in Alzheimer disease. *Arch Neurol* **67**, 677-685.
- [14] Kauwe JS, Bailey MH, Ridge PG, Perry R, Wadsworth ME, Hoyt KL, Staley LA, Karch CM, Harari O, Cruchaga C, Ainscough BJ, Bales K, Pickering EH, Bertelsen S, Fagan AM, Holtzman DM, Morris JC, Goate AM (2014) Genome-wide association study of CSF levels of 59 Alzheimer's disease candidate proteins: Significant associations with proteins involved in amyloid processing and inflammation. *PLoS Genet* **10**, e1004758.
- [15] Cruchaga C, Kauwe JS, Harari O, Jin SC, Cai Y, Karch CM, Benitez BA, Jeng AT, Skorupa T, Carrell D, et al. (2013) GWAS of cerebrospinal fluid tau levels identifies risk variants for Alzheimer's disease. *Neuron* **78**, 256-268.
- [16] Barral S, Bird T, Goate A, Farlow MR, Diaz-Arrastia R, Bennett DA, Graff-Radford N, Boeve BF, Sweet RA, Stern Y, Wilson RS, Foroud T, Ott J, Mayeux R, Green R, Kowall N, Farrer L, Williamson J, Santana V, Schmechel D, Gaskell P, Ghetti B, Farlow MR, Faber K, Prentice H, Horner K, Growdon JH, Blacker D, Tanzi RE, Hyman BT, Boeve B, Kuntz K, Norgaard L, Larson N, Kistler D, Parfitt F, Haddow J, Silverman J, Beeri MS, Sano M, Wang J, Lally R, Johnson N, Mesulam M, Weintraub S, Bigio E, Kaye J, Kramer P, Payne-Murphy J, Bennett D, Jacobs H, Chang JS, Arends D, Harrell L, Bartzokis G, Cummings J, Lu PH, Toland U, Smith C, Brickhouse A, Trojanowski J, Van Deerlin V, Wood EM, DeKosky S, Sweet R, Weamer E, Chui H, Varpetian A, Diaz-Arrastia R, Rosenberg R, Davis B, Bird T, Schellenberg GD, Raskind M, Rumbaugh M, Nickel K, Goate A, Morris J, Norton J, Levitch D, Grant B, Coats M, Levey A, Rosen A, Anosike E (2012) Genotype patterns at PICALM, CR1, BIN1, CLU, and APOE genes are associated with episodic memory. *Neurology* **78**, 1464-1471.
- [17] Barral S, Cosentino S, Christensen K, Newman AB, Perls TT, Province MA, Mayeux R (2014) Common genetic variants on 6q24 associated with exceptional episodic memory performance in the elderly. *JAMA Neurol* **71**, 1514-1519.
- [18] Beecham GW, Hamilton K, Naj AC, Martin ER, Huentelman M, Myers AJ, Corneveaux JJ, Hardy J, Vonsattel JP, Younkin SG, Bennett DA, De Jager PL, Larson EB, Crane PK, Kamboh MI, Kofler JK, Mash DC, Duque L, Gilbert JR, Gwirtsman H, Buxbaum JD, Kramer P, Dickson DW, Farrer LA, Frosch MP, Ghetti B, Haines JL, Hyman BT, Kukull WA, Mayeux RP, Pericak-Vance MA, Schneider JA, Trojanowski JQ, Reiman EM, Schellenberg GD, Montine TJ (2014) Genome-wide association meta-analysis of neuropathologic features of Alzheimer's disease and related dementias. *PLoS Genet* **10**, e1004606.
- [19] Jansen IE, Savage JE, Watanabe K, Bryois J, Williams DM, Steinberg S, Sealock J, Karlsson IK, Hägg S, Athanasiu L, Voyle N, Proitsi P, Witoelar A, Stringer S, Aarsland D, Almdahl IS, Andersen F, Bergh S, Bettella F, Bjornsson S, Brækhus A, Bråthen G, de Leeuw C, Desikan RS, Djurovic S, Dumitrescu L, Fladby T, Hohman TJ, Jonsson P V, Kiddle SJ, Rongve A, Saltvedt I, Sando SB, Selbæk G, Shoi M, Skene NG, Snaedal J, Stordal E, Ulstein ID, Wang Y, White LR, Hardy J, Hjerling-Leffler J, Sullivan PF, van der Flier WM, Dobson R, Davis LK, Stefansson H, Stefansson K, Pedersen NL, Ripke S, Andreassen OA, Posthuma D (2019) Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat Genet* **51**, 404-413.
- [20] Cukier HN, Kunkle BW, Vardarajan BN, Rolati S, Hamilton-Nelson KL, Kohli MA, Whitehead PL, Dombroski BA, Van Booven D, Lang R, Dykxhoorn DM, Farrer LA, Cuccaro ML, Vance JM, Gilbert JR, Beecham GW, Martin ER, Carney RM, Mayeux R, Schellenberg GD, Byrd GS, Haines JL, Pericak-Vance MA (2016) ABCA7 frameshift deletion associated with Alzheimer disease in African Americans. *Neurol Genet* **2**, e79.
- [21] Mez J, Chung J, Jun G, Kriegel J, Bourlas AP, Sherva R, Logue MW, Barnes LL, Bennett DA, Buxbaum JD, Byrd GS, Crane PK, Ertekin-Taner N, Evans D, Fallin MD, Foroud T, Goate A, Graff-Radford NR, Hall KS, Kamboh MI, Kukull WA, Larson EB, Manly JJ, Haines JL, Mayeux R, Pericak-Vance MA, Schellenberg GD, Lunetta KL, Farrer LA (2017) Two novel loci, COBL and SLC10A2, for Alzheimer's disease in African Americans. *Alzheimers Dement* **13**, 119-129.
- [22] Hirano A, Ohara T, Takahashi A, Aoki M, Fuyuno Y, Ashikawa K, Morihara T, Takeda M, Kamino K, Oshima E, Okahisa Y, Shibata N, Arai H, Akatsu H, Ikeda M, Iwata N, Ninomiya T, Monji A, Kitazono T, Kiyohara Y, Kubo M, Kanba S (2015) A genome-wide association study of late-onset Alzheimer's disease in a Japanese population. *Psychiatr Genet* **25**, 139-146.
- [23] Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, McMahon A, Morales J, Mountjoy E, Sollis E, Suveges D, Vrousou O, Whetzel PL, Amode R, Guillen JA, Riat HS, Trevanion SJ, Hall P, Junkins H, Flicek P, Burdett T, Hindorf LA, Cunningham F, Parkinson H (2019) The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res* **47**, D1005-D1012.
- [24] Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K (2001) dbSNP: The NCBI database of genetic variation. *Nucleic Acids Res* **29**, 308-311.
- [25] Yates AD, Achuthan P, Akanni W, Allen J, Allen J, Alvarez-Jarreta J, Amode MR, Armean IM, Azov AG, Bennett R, Bhai J, Billis K, Boddu S, Marugán JC, Cummins C, Davidson C, Dodiya K, Fatima R, Gall A, Giron CG, Gil L, Grego T, Haggerty L, Haskell E, Hourlier T, Izuogu OG, Janacek SH, Juettemann T, Kay M, Lavidas I, Le T, Lemos D, Martinez JG, Maurel T, McDowall M, McMahon A, Mohanan S, Moore B, Nuhn M, Ohed DN, Parker A, Parton A, Patricio M, Sakthivel MP, Abdul Salam AI, Schmitt BM, Schuilenburg H, Sheppard D, Sycheva M, Szuba M, Taylor K, Thormann A, Threadgold G, Vullo A, Walts B, Winterbottom A, Zadissa A, Chakiachvili M, Flint B, Frankish A, Hunt SE, Iisley G, Kostadima M, Langridge N, Loveland

- JE, Martin FJ, Morales J, Mudge JM, Muffato M, Perry E, Ruffier M, Trevanion SJ, Cunningham F, Howe KL, Zerbino DR, Flicek P (2020) Ensembl 2020. *Nucleic Acids Res* **48**, D682-D688.
- [26] NIAGADS Alzheimer's GenomicsDB: A resource for exploring Alzheimer's disease genetic and genomic knowledge. <https://www.niagads.org/genomics/>, Last updated June 26, 2021, Accessed on November 01, 2021.
- [27] Frankish A, Diekhans M, Ferreira A-M, Johnson R, Jungreis I, Loveland J, Mudge JM, Sisu C, Wright J, Armstrong J, Barnes I, Berry A, Bignell A, Carbonell Sala S, Chrast J, Cunningham F, Di Domenico T, Donaldson S, Fiddes IT, García Girón C, Gonzalez JM, Grego T, Hardy M, Hourlier T, Hunt T, Izuogu OG, Lagarde J, Martin FJ, Martínez L, Mohanan S, Muir P, Navarro FCP, Parker A, Pei B, Pozo F, Ruffier M, Schmitt BM, Stapleton E, Suner M-M, Sycheva I, Uszczynska-Ratajczak B, Xu J, Yates A, Zerbino D, Zhang Y, Aken B, Choudhary JS, Gerstein M, Guigó R, Hubbard TJP, Kellis M, Paten B, Reymond A, Tress ML, Flicek P (2019) GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res* **47**, D766-D773.
- [28] Braschi B, Denny P, Gray K, Jones T, Seal R, Tweedie S, Yates B, Bruford E (2019) Genenames.org: The HGNC and VGNC resources in 2019. *Nucleic Acids Res* **47**, D786-D792.
- [29] Amlie-Wolf A, Tang M, Mlynarski EE, Kuksa PP, Valladares O, Katanic Z, Tsuang D, Brown CD, Schellenberg GD, Wang L-S (2018) INFERNO: Inferring the molecular mechanisms of noncoding genetic variants. *Nucleic Acids Res* **46**, 8740-8753.
- [30] Kuksa PP, Lee C-Y, Amlie-Wolf A, Gangadharan P, Mlynarski EE, Chou Y-F, Lin H-J, Issen H, Greenfest-Allen E, Valladares O, Leung YY, Wang L-S (2020) SparkINFERNO: A scalable high-throughput pipeline for inferring molecular mechanisms of non-coding genetic variants. *Bioinformatics* **36**, 3879-3881.
- [31] Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR (2015) A global reference for human genetic variation. *Nature* **526**, 68-74.
- [32] Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, Hasz R, Walters G, Garcia F, Young N, Foster B, Moser M, Karasik E, Gillard B, Ramsey K, Sullivan S, Bridge J, Magazine H, Syron J, Fleming J, Siminoff L, Traino H, Mosavel M, Barker L, Jewell S, Rohrer D, Maxim D, Filkins D, Harbach P, Cortadillo E, Berghuis B, Turner L, Hudson E, Feenstra K, Sobin L, Robb J, Branton P, Korzeniewski G, Shive C, Tabor D, Qi L, Groch K, Nampally S, Buia S, Zimmerman A, Smith A, Burges R, Robinson K, Valentino K, Bradbury D, Cosentino M, Diaz-Mayoral N, Kennedy M, Engel T, Williams P, Erickson K, Ardlie K, Winckler W, Getz G, DeLuca D, Daniel MacArthur, Kellis M, Thomson A, Young T, Gelfand E, Donovan M, Meng Y, Grant G, Mash D, Marcus Y, Basile M, Liu J, Zhu J, Tu Z, Cox NJ, Nicolae DL, Gamazon ER, Im HK, Konkashbaev A, Pritchard J, Stevens M, Flutre T, Wen X, Dermitzakis ET, Lappalainen T, Guigo R, Monlong J, Sammeth M, Koller D, Battle A, Mostafavi S, McCarthy M, Rivas M, Maller J, Rusyn I, Nobel A, Wright F, Shabalina A, Feolo M, Sharopova N, Sturcke A, Paschal J, Anderson JM, Wilder EL, Derr LK, Green ED, Struwing JP, Temple G, Volpi S, Boyer JT, Thomson EJ, Guyer MS, Ng C, Abdallah A, Colantuoni D, Insel TR, Koester SE, Roger Little A, Bender PK, Lehner T, Yao Y, Compton CC, Vaught JB, Sawyer S, Lockhart NC, Demchok J, Moore HF (2013) The Genotype-Tissue Expression (GTEx) project. *Nat Genet* **45**, 580-585.
- [33] Kundaje A, Meuleman W, Ernst J, Bilienky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, Ziller MJ, Amin V, Whitaker JW, Schultz MD, Ward LD, Sarkar A, Quon G, Sandstrom RS, Eaton ML, Wu Y-C, Pfenning AR, Wang X, Claussnitzer M, Liu Y, Coarfa C, Harris RA, Shoresh N, Epstein CB, Gjoneska E, Leung D, Xie W, Hawkins RD, Lister R, Hong C, Gascard P, Mungall AJ, Moore R, Chuah E, Tam A, Canfield TK, Hansen RS, Kaul R, Sabo PJ, Bansal MS, Carles A, Dixon JR, Farh K-H, Feizi S, Karlic R, Kim A-R, Kulkarni A, Li D, Lowdon R, Elliott G, Mercer TR, Neph SJ, Onuchic V, Polak P, Rajagopal N, Ray P, Sallari RC, Siebenthal KT, Sinnott-Armstrong NA, Stevens M, Thurman RE, Wu J, Zhang B, Zhou X, Beaudet AE, Boyer LA, De Jager PL, Farnham PJ, Fisher SJ, Haussler D, Jones SJM, Li W, Marra MA, McManus MT, Sunyaev S, Thomson JA, Tlsty TD, Tsai L-H, Wang W, Waterland RA, Zhang MQ, Chadwick LH, Bernstein BE, Costello JF, Ecker JR, Hirst M, Meissner A, Milosavljevic A, Ren B, Stamatoyannopoulos JA, Wang T, Kellis M (2015) Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317-330.
- [34] Ernst J, Kellis M (2012) ChromHMM: Automating chromatin-state discovery and characterization. *Nat Methods* **9**, 215-216.
- [35] Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, Chen Y, Zhao X, Schmid C, Suzuki T, Ntini E, Arner E, Valen E, Li K, Schwarzfischer L, Glatz D, Raithel J, Lilje B, Rapin N, Bagger FO, Jorgensen M, Andersen PR, Bertin N, Rackham O, Burroughs AM, Bailie JK, Ishizu Y, Shimizu Y, Furuhashi E, Maeda S, Negishi Y, Mungall CJ, Meehan TF, Lassmann T, Itoh M, Kawaji H, Kondo N, Kawai J, Lennartsson A, Daub CO, Heutink P, Hume DA, Jensen TH, Suzuki H, Hayashizaki Y, Muller F, Forrest ARR, Carninci P, Rehli M, Sandelin A (2014) An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455-461.
- [36] Bertram L, McQueen MB, Mullin K, Blacker D, Tanzi RE (2007) Systematic meta-analyses of Alzheimer disease genetic association studies: The AlzGene database. *Nat Genet* **39**, 17-23.
- [37] Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* **6**, 80-92.
- [38] Butkiewicz M, Blue EE, Leung YY, Jian X, Marcora E, Renton AE, Kuzma A, Wang LS, Koboldt DC, Haines JL, Bush WS (2018) Functional annotation of genomic variants in studies of late-onset Alzheimer's disease. *Bioinformatics* **34**, 2724-2731.
- [39] Navarro Gonzalez J, Zweig AS, Speir ML, Schmelter D, Rosenbloom KR, Raney BJ, Powell CC, Nassar LR, Maulding ND, Lee CM, Lee BT, Hinrichs AS, Fyfe AC, Fernandes JD, Diekhans M, Clawson H, Casper J, Benet-Pagès A, Barber GP, Haussler D, Kuhn RM, Haussler M, Kent WJ (2021) The UCSC Genome Browser database: 2021 update. *Nucleic Acids Res* **49**, D1046-D1057.
- [40] Koch L (2020) Exploring human genomic diversity with gnomAD. *Nat Rev Genet* **21**, 448.
- [41] McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, Flicek P, Cunningham F (2016) The Ensembl Variant Effect Predictor. *Genome Biol* **17**, 122.

- [42] Reitz C, Jun G, Naj A, Rajbhandary R, Vardarajan BN, Wang L-S, Valladares O, Lin C-F, Larson EB, Graff-Radford NR, Evans D, De Jager PL, Crane PK, Buxbaum JD, Murrell JR, Raj T, Ertekin-Taner N, Logue M, Baldwin CT, Green RC, Barnes LL, Cantwell LB, Fallin MD, Go RCP, Griffith P, Obisesan TO, Manly JJ, Lunetta KL, Kamboh MI, Lopez OL, Bennett DA, Hendrie H, Hall KS, Goate AM, Byrd GS, Kukull WA, Foroud TM, Haines JL, Farrer LA, Pericak-Vance MA, Schellenberg GD, Mayeux R, Alzheimer Disease Genetics Consortium (2013) Variants in the ATP-binding cassette transporter (ABCA7), apolipoprotein E  $\epsilon$ 4, and the risk of late-onset Alzheimer disease in African Americans. *JAMA* **309**, 1483-1492.
- [43] Thomas RS, Henson A, Gerrish A, Jones L, Williams J, Kidd EJ (2016) Decreasing the expression of PICALM reduces endocytosis and the activity of  $\beta$ -secretase: Implications for Alzheimer's disease. *BMC Neurosci* **17**, 50.
- [44] Hodes RJ, Buckholtz N (2016) Accelerating Medicines Partnership: Alzheimer's Disease (AMP-AD) knowledge portal aids alzheimer's drug discovery through open data sharing. *Expert Opin Ther Targets* **20**, 389-391.
- [45] Sieberts SK, Perumal TM, Carrasquillo MM, Allen M, Reddy JS, Hoffman GE, Dang KK, Calley J, Ebert PJ, Eddy J, Wang X, Greenwood AK, Mostafavi S, CommonMind Consortium (CMC), The AMP-AD Consortium, Omberg L, Peters MA, Logsdon BA, De Jager PL, Ertekin-Taner N, Mangravite LM (2020) Large eQTL meta-analysis reveals differing patterns between cerebral cortical and cerebellar brain regions. *Sci Data* **7**, 340.
- [46] Reymond N, Imbert A-M, Devillard E, Fabre S, Chabannon C, Xerri L, Farnarier C, Cantoni C, Bottino C, Moretta A, Dubreuil P, Lopez M (2004) DNAM-1 and PVR regulate monocyte migration through endothelial junctions. *J Exp Med* **199**, 1331-1341.
- [47] Kehoe PG (2018) The coming of age of the angiotensin hypothesis in Alzheimer's disease: Progress toward disease prevention and treatment? *J Alzheimers Dis* **62**, 1443-1466.
- [48] Heneka MT, Golenbock DT, Latz E (2015) Innate immunity in Alzheimer's disease. *Nat Immunol* **16**, 229-236.
- [49] Jiang C, Li G, Huang P, Liu Z, Zhao B (2017) The gut microbiota and Alzheimer's disease. *J Alzheimers Dis* **58**, 1-15.
- [50] Loskutova N, Honea RA, Vidoni ED, Brooks WM, Burns JM (2009) Bone density and brain atrophy in early Alzheimer's disease. *J Alzheimers Dis* **18**, 777-785.
- [51] Wan Y-W, Al-Ouran R, Mangleburg CG, Perumal TM, Lee T V, Allison K, Swarup V, Funk CC, Gaiteri C, Allen M, Wang M, Neuner SM, Kaczorowski CC, Philip VM, Howell GR, Martini-Stoica H, Zheng H, Mei H, Zhong X, Kim JW, Dawson VL, Dawson TM, Pao P-C, Tsai L-H, Haure-Mirande J-V, Ehrlich ME, Chakrabarty P, Levites Y, Wang X, Dammer EB, Srivastava G, Mukherjee S, Sieberts SK, Omberg L, Dang KD, Eddy JA, Snyder P, Chae Y, Amberkar S, Wei W, Hide W, Preuss C, Ergun A, Ebert PJ, Airey DC, Mostafavi S, Yu L, Klein H-U, Accelerating Medicines Partnership-Alzheimer's Disease Consortium, Carter GW, Collier DA, Golde TE, Levey AI, Bennett DA, Estrada K, Townsend TM, Zhang B, Schadt E, De Jager PL, Price ND, Ertekin-Taner N, Liu Z, Shulman JM, Mangravite LM, Logsdon BA (2020) Meta-analysis of the Alzheimer's disease human brain transcriptome and functional dissection in mouse models. *Cell Rep* **32**, 107908.
- [52] Amlie-Wolf A, Tang M, Way J, Dombroski B, Jiang M, Vrettos N, Chou Y-F, Zhao Y, Kuzma A, Mlynarski EE, Leung YY, Brown CD, Wang L-S, Schellenberg GD (2019) Inferring the molecular mechanisms of non-coding Alzheimer's disease-associated genetic variants. *J Alzheimers Dis* **72**, 301-318.
- [53] Kikuchi M, Hara N, Hasegawa M, Miyashita A, Kuwano R, Ikeuchi T, Nakaya A (2019) Enhancer variants associated with Alzheimer's disease affect gene expression via chromatin looping. *BMC Med Genomics* **12**, 128.
- [54] Sims R, van der Lee SJ, Naj AC, Bellenguez C, Badarinarayan N, Jakobsdottir J, Kunkle BW, Boland A, Raybould R, Bis JC, et al. (2017) Rare coding variants in PLCG2, ABI3, and TREM2 implicate microglial-mediated innate immunity in Alzheimer's disease. *Nat Genet* **49**, 1373-1384.
- [55] Kundu P, Torres ERS, Stagaman K, Kasschau K, Okhovat M, Holden S, Ward S, Nevenon KA, Davis BA, Saito T, Saido TC, Carbone L, Sharpton TJ, Raber J (2021) Integrated analysis of behavioral, epigenetic, and gut microbiome analyses in AppNL-G-F, AppNL-F, and wild type mice. *Sci Rep* **11**, 4678.
- [56] Kuzma A, Valladares O, Cweibel R, Greenfest-Allen E, Childress DM, Malamon J, Gangadharan P, Zhao Y, Qu L, Leung YY, Naj AC, Stoeckert Jr. CJ, Schellenberg GD, Wang L-S (2016) NIAGADS: The NIA Genetics of Alzheimer's Disease Data Storage Site. *Alzheimers Dement* **12**, 1200-1203.