Alzheimer's disease variant portal (ADVP): a catalog of genetic findings for Alzheimer's disease

Pavel P. Kuksa[1,2,*], Chia-Lun Liu[1,2,*], Wei Fu[1,2,*], Liming Qu[1,2], Yi Zhao[1,2], Zivadin Katanic[1,2], Amanda B Kuzma[1,2], Pei-Chuan Ho[4], Kai-Teh Tzeng[4], Otto Valladares[1,2], Shin-Yi Chou[1,2,4], Adam C Naj[1,3], Gerard D Schellenberg[1,2], Li-San Wang[1,2,^] and Yuk Yee Leung[1,2,^]

[1]Penn Neurodegeneration Genomics Center, [2] Department of Pathology and Laboratory Medicine, [3]Biostatistics and Epidemiology, University of Pennsylvania Perelman School of Medicine, Philadelphia PA, USA, [4] Department of Economics, Lehigh University, Bethlehem, PA, USA

** co-first author*

^co-corresponding author

Abstract

Background:

Alzheimer's disease (AD) genetic findings span progressively larger genome-wide association studies (GWASs) for various outcomes and populations. These genetic findings are obtained from a single GWAS, joint- or meta- analyses of multiple GWAS datasets. However, no single resource provides harmonized and searchable information on all AD genetic associations obtained from these analyses, nor linking the identified genetic variants and reported genes with other supporting functional genomic evidence.

Methods:

We created the Alzheimer's Disease Variant Portal (ADVP), which provides unified access to a uniquely extensive collection of high-quality GWAS association results for AD. Records in ADVP are curated from the genome-wide significant and suggestive loci reported in AD genetics literature. ADVP contains curated results from all AD GWAS publications by Alzheimer's Disease Genetics Consortium (ADGC) since 2009 and AD

GWAS publications identified from other public catalogs (GWAS catalog). Genetic association information was systematically extracted from these publications, harmonized, and organized into three types of tables. These tables included structured publication, variant, and association categories to ensure consistent representation of all AD genetic findings. All extracted AD genetic associations were further annotated and integrated with NIAGADS Alzheimer's Genomics DB in order to provide extensive biological and functional genomics annotations.

Results:

Currently, ADVP contains 6,990 AD-association records curated from >200 AD GWAS publications corresponding to >900 unique genomic loci and >1,800 unique genetic variants. The ADVP collection contains genetic findings from >80 cohorts and across various populations, including Caucasians, Hispanics, African-Americans, and Asians. Of all the association records, 46% are disease-risk, 13% are related to expression quantitative trait analyses, and 27% are related to AD endophenotypes and neuropathology. ADVP web interface allows accessing AD association records by individual variants, genes, publications, genomic regions of interest, and genome-wide interactive variant views. ADVP is integrated with the NIAGADS Alzheimer's Genomics Database. Researchers can explore additional biological annotations at the genetic variant or gene level and view cross-reference functional genomics evidence provided by other public resources.

Conclusions:

ADVP is the largest, most up-to-date, and comprehensive literature-derived collection of AD genetic associations. All records have been systematically curated, harmonized, and comprehensively annotated. ADVP is freely accessible at https://advp.niagads.org/.

Introduction

Alzheimer's disease (AD) is a devastating neurological disorder affecting millions of people worldwide and is the most common cause of dementia (Association, 2019). There are no approved drugs that can slow or treat the disease. The disease is complex and is highly heritable (Gatz et al., 1997). The strongest known genetic risk factor for AD is the ε4 allele of the Apolipoprotein E gene (*APOE* ε4) (Corder et al., 1993; Genin et al., 2011), but more than one-third of AD cases do not carry any APOE ε4 alleles. Large-scale genome-wide association studies (GWASs) have been performed to find more genetic risk factors. These led to the discovery of additional common genetic loci associated with the late-onset AD (LOAD) (Harold et al., 2009; Hollingworth et al., 2011; Lambert et al., 2009; Naj et al., 2011; Seshadri et al., 2010). Yet, the identification of genetic contributors to LOAD remains a challenge as LOAD is likely caused by multiple low penetrance genetic variants (Naj & Schellenberg, 2017), 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 (Lambert et al., 2013), and five more loci using more than 30,000 samples in 2019 (Kunkle et al., 2019).

In addition to GWAS studies focused on association with disease risk, recently many genetics studies have focused on related phenotypes including, e.g., neuroimaging biomarkers (Biffi et al., 2010), circulating biomarkers in (Cruchaga et al., 2013; Kauwe et al., 2014), cognitive decline (Barral et al., 2012, 2014) neuropathology (Beecham et

al., 2014), family history (Jansen et al., 2019). GWAS on Hispanic, African-American, Asian, and other minority populations also led to new variants not observed in Caucasians (Cukier et al., 2016; Hirano et al., 2015; Mez et al., 2017) In order to help researchers 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 genetic association results (both genome-wide significant and suggestive) from all major GWAS studies published by ADGC from 2009 to 2019 and other AD GWAS studies identified from publicly available catalogs (GWAS catalog, Buniello et al., 2019). All the collected data from each of the association studies are made publicly available on a continuously updated and freely accessible 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, >900 genomic loci curated from >125 AD publications categorized into nine harmonized phenotype categories. All AD associations in ADVP are annotated with genomic and functional information. Comprehensive biological annotations are available via integration with the NIAGADS Alzheimer's Disease Genomics database (GenomicsDB, 2020). ADVP will serve as an invaluable resource for the research community to explore and decipher the genetic architecture of AD and other neurodegenerative diseases.

Methods

An overview of the ADVP study design is shown in Figure 1.

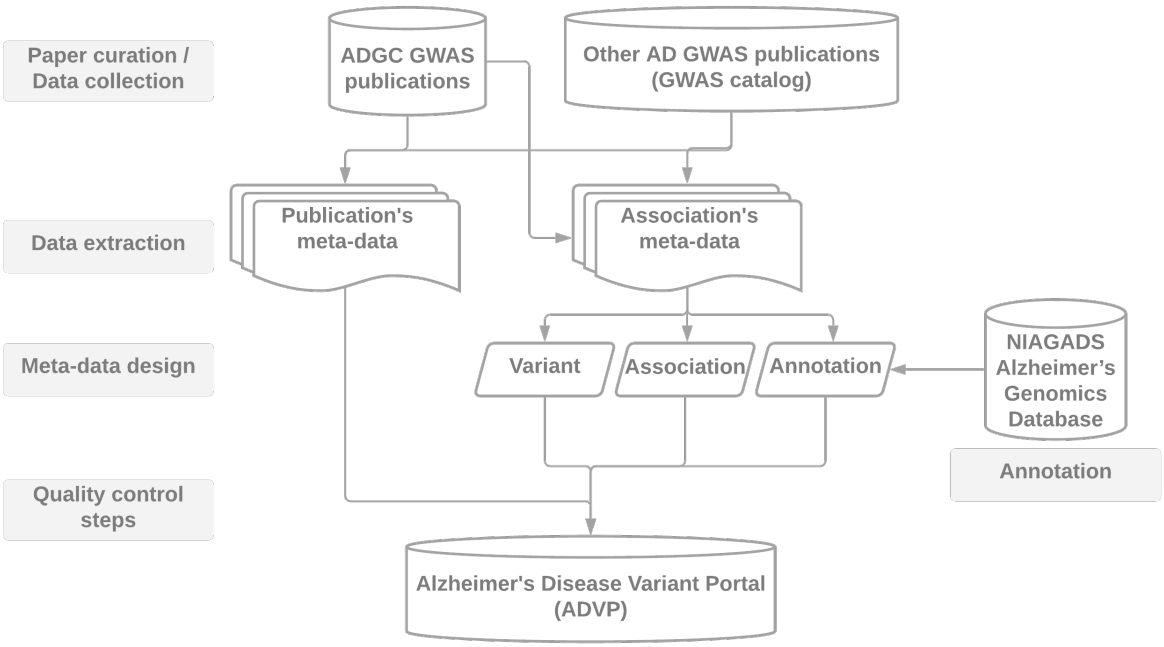


Figure 1. ADVP study design. AD GWAS publications are first collected (Section **"Data collection"**), genetic variant and association data are then systematically extracted (Section "**Data extraction**"), harmonized (Section "**Meta-data design**"), annotated (Section "**Annotation**"), subjected to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

Collection and curation of AD-related GWAS publications (Data collection)

This ADVP V1.0 release consists of curated and harmonized genetic associations from the genome-wide significant and suggestive loci collected from AD genetic studies conducted primarily by the ADGC. All AD GWAS publications by ADGC (2009-2019, http://www.adgenetics.org) and all other AD GWAS studies in GWAS catalog (Buniello et al., 2019) (MeSH D000544, curation date: Dec 2019) were included. All publications (total N=205; ADGC: N=134; Citations from ADGC: N=20; GWAS catalog: N=51) fulfilling the above search criteria were first screened to identify publications reporting GWAS findings. All reported genetic associations in the main text (table format) were systematically extracted. We curated from the 125 publications that met the above criteria (https://advp.niagads.org/publications). Supplementary Table S1 provides

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

Extraction of genetic variants and associations from publications (Data extraction)

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 ADVP data schema (see Section "**Meta-data design**" for details about the 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 a standardized template document using the corresponding meta-data schema. The completed document for all curated publications is composed of the three predefined worksheets:

1. The publication's meta-data
2. Association meta-data

Lastly, the document is parsed by customized scripts to normalize, validate, annotate, and store the publication, variant, and association data in the relational database (https://dl.acm.org/doi/book/10.5555/560480). Collected AD variants and association records are further integrated with the NIAGADS Alzheimer's Genomics Database (GenomicsDB, 2020) providing, comprehensive genomic annotation and functional genomic information.

Meta-data schema for systematic curation and harmonization of genetic associations (Meta-data design)

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 interface (https://www.ncbi.nlm.nih.gov/books/NBK179288/) based on the publication's PubMed

identifier (PMID). For each publication, we record its PMID, PubMed Central identifier (PMCID), first and last authors, journal, and publication year. We also store the abstract, article URLs, as well as information on the curated tables for each article in the Publication meta-data (Figure 1).

Association meta-data

ADVP association meta-data consists of 28 data fields, of which 19 are extracted directly from the paper contents and nine fields are the additional, harmonized (based on extracted original information) and programmatically generated fields. Altogether, for each AD association, association meta-data provide 1) variant information (see Section **Description of Variants**); 2) association information (see Section **Description of association records**); 3) annotation information (see Section **Annotation of genetic variants and associations**). For a detailed explanation of these curated and harmonized/derived data fields, see Supplementary Table S2.

Description of Variants

Each genetic variant in ADVP is described using dbSNP rsID, genomic coordinates (chromosome: basepair), genomic reference, and alternative alleles. Both the values reported in the publication (if available) and the values derived from the reference databases such as dbSNP (Sherry et al., 2001) and NIAGADS Alzheimer's Genomics database (GenomicsDB, 2020), are included in the variant description. Genomic location in the current version of ADVP is stored using GRCh37/hg19 reference genome build as the majority of GWAS publications conducted analyses using GRCh37/hg19.

For quality assurance, reported rsID, coordinate, and allele information for each variant were cross-checked against dbSNP b151 ((GenomicsDB, 2020; Sherry et al., 2001)) and referenced with 1000 genome data (Auton et al., 2015) to help resolve reported alleles and complete any missing variation information (e.g., genomic coordinates, allele information).

Description of association records

Association information was systematically extracted from each source table and recorded as a part of ADVP association record. They were further recoded and categorized, so that association records are described consistently across publications/studies. For each reported association ADVP first collected a pre-defined set of data attributes commonly reported by genetic association studies (See "Extracted" columns under Supplementary Table S2). These include *p*-value and statistics related to the effect of the genetic variant (regression beta coefficients and variance, odds ratios, confidence intervals), reported effect allele, and its frequency in the studied population.

In addition to the information directly extracted from publications, each association in ADVP is described with the nine meta-information data fields:

1) "Record type": association record type is set based on whether the reported association is for a single SNP ("SNP-level"), a single gene ("Gene-level"), SNP interactions ("Interaction (SNP)", or gene interactions ("Interaction (Gene)").
2) "Population": study population information was first copied from the publication ("Population (detailed)"). Then, the reported population information was further mapped using standard population vocabulary to normalize population information across studies ("Population" column).
3) "Cohort" - Full cohort names ("Cohort (detailed)") were mapped to consortium names whenever available to obtain normalized cohort information ("Cohort"). For example, if the cohort is one of the ADGC GWAS cohorts (see Supplementary Table S3), ADVP appends "ADGC" to the data entry.
4) "Sample size": Original sample size (number of cases, number of controls when available/applicable) were recorded. ADVP V1.0 webserver only reports the total number.
5) "Subset analyzed": description of the subset of samples used in association analysis. When a subset of samples was used to perform association analysis

(e.g., “e2/e4”, “e4 carriers only”, “Female only”), this field records description of the subset as described in the publication.

6) “Phenotype”: the outcome variable (i.e., phenotype/trait) of the association analysis. Original outcome (“Phenotype (detailed)”) was assigned to one of the nine categories including “AD”, ADRD”, “Cognitive”, “Expression”, “Fluid biomarker”, “Imaging”, “Neuropathology”, Non-ADRD”, and “Other” (representing “Age of onset” or “AD survival”).
7) “Association Type”: the type of association analysis. Based on the “Phenotype” column, we classified each association test into six categories: “Age at onset (AAO)/ Survival”, “Cross phenotype”, “Disease-risk”, “Endophenotype”, expression quantitative trait locus “eQTL”, “Pleiotropy”.
8) “Stage”: the stage of the analysis as described in the publication, e.g., “Stage n” (n=1,2,3), “Discovery”, “Validation”, “Meta-analysis”. The stage information is reported as given in the paper if this information was available. If the stage information was not explicitly provided in the text, we derived the stage information as follows (Supplementary Figure S1). If the association analysis was done using a single cohort, the “Stage” of the record was set as: “Stage n” (n=1,2,3 or others); “Discovery”, “Replication/Validation”. “Discovery” was used if the paper was the first to report such findings using the specific combination of cohort + phenotype information, otherwise, the stage was set as “Replication/Validation”. If the association analysis was done using multiple cohorts, the stage was set as “Meta-analysis” if the analyses were performed using methods such as “inverse-variance weighting”, “fixed effects”, “random effects” model or METAL R package; if not, the stage was set to “Joint-analysis”.
9) “Imputation”: imputation panel information. The imputation panel version, software tool and version were mapped to broader categories such as 1000 Genome project (“1000G”), International HapMap Project (“HapMap”) or Haplotype Reference Consortium (“HRC”).

Note, population, Cohort, and Phenotype information are displayed in ADVP using both the original (reported) and the derived, harmonized data columns.

Functional genomics evidence for genetic variants and associations (Annotation)

All variants and association in ADVP were systematically annotated with genomic context (closest upstream/downstream genes), genomic element (promoter, UTR, intron, exon, intergenic, repeat), functional information (variant most severe consequence), and cross-reference to NIAGADS Alzheimer's Genomics database (GenomicsDB, 2020).

ADVP reports the genomic context of each genetic variant via multiple data fields: 1) "Locus" field – records the gene name as reported in the publication; 2) "Nearest gene" – contains the name of the gene closest to the variant as well as the distance to the gene (in base pairs (bp)) in upstream (+) or downstream (-) orientation. The nearest genes are identified by using GENCODE v34 (Frankish et al., 2019) protein-coding gene annotations. Both Ensembl ID (Yates et al., 2020) and HGNC (Braschi et al., 2019) gene symbols are reported for each genetic variant. For each ADVP variant, the co-localized genomic element (e.g., promoter, intron, UTR, repeat, intergenic) is reported based on the genomic partition information (Amlie-Wolf et al., 2018; Kuksa et al., 2020).

Variant and association data verification (Quality control steps)

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

1) We ensured records are not double-counted/re-reported across studies. Each association record in ADVP is uniquely identified by a combination of reported gene/SNP/interaction name, cohort/analyzed subset, the model used, phenotype, and association *p*-value and effect size.
2) We cross-checked recorded positional information (chromosome: base pair), rsID, and allele information against reference databases including, dbSNP

(Sherry et al., 2001) and NIAGADS Alzheimer's Genomics database (GenomicsDB, 2020) to ensure variant information is correct.

3) We identified and removed records solely representing variants annotated by publicly available functional resources such as GTEx (Lonsdale et al., 2013).

ADVP front-end and back-end architecture and implementation

ADVP is designed with ease of update and modularity in mind. Contents of ADVP are derived from collection, curation, harmonization, processing and integration of AD-related publications and reported genetic associations using a meta-table scheme (see Section **Meta-data curation**). The ADVP web server runs on Amazon Web Services (AWS) cloud computing instance (m5.4xlarge) using MySQL (Widenius et al., 2002) relational database management system as a back-end and a PHP/JQuery-based web front-end. All the publication, variant and association information stored in ADVP relational database is organized into multiple tables (Figure 1). The web front-end provides multiple data views for publications, genes, variants, and association records (Figure 4).

Results

ADVP is more comprehensive than the NHGRI-EBI GWAS catalog (Buniello et al., 2019) (Table 1) in terms of the number of curated AD-related associations and publications, and more recent than another major database AlzGene (Bertram et al., 2007). In order to focus on association findings with the highest confidence, we decided to focus on large-scale association studies at the genomic level, with the majority of studies included in ADVP (65%) reporting associations reaching genome-wide significance, the gold standard for GWAS discoveries. Furthermore, ADVP collected extensive meta-data, including consortiums and cohorts, which were not available in the other two databases and are important for relating the results reported across publications. Finally, ADVP provides convenient links for researchers to explore

biological significance via an annotation in NIAGADS Alzheimer's Genomics database (GenomicsDB, 2020) (Figure 1).

Table 1: Comparison between ADVP and existing AD genetics database (AlzGene and GWAS Catalog).

Features	**AlzGene** (Bertram et al., 2007)	**GWAS Catalog** (Buniello et al., 2019)	**ADVP**
Focus on AD	Yes	No	Yes
Types of genetic associations recorded / included	Variant only	Variant only	Variant, Gene, SNP-SNP and gene-gene interactions
Included associations (*p*-value cutoff)	Genome-wide significant only ($<5*10^{-8}$)	Genome-wide significant only ($<5*10^{-8}$)	Both genome-wide significant and suggestive associations
Number of curated AD GWAS publications (publication years)	41 (1998-2011)	69 (2007-current)	Initial set: 205; Curated set: 125 (2009-current)
Number of curated records/associations	680 genes (2,611 variants)	1,532 records (1,155 variants)	6,605 AD, 6,990 total association records (1,825 variants)
Genomic and functional genomic annotations	No	No	NIAGADS Alzheimer's Genomics database (GenomicsDB, 2020) annotations
Detailed / extensive meta-data	Population, cohort, sample size, phenotype	Population, sample size, stage, phenotype	Record type (SNP/gene-based), population, cohort, sample size, subset analyzed,

			phenotype, association type, stage, imputation
Last update date	2011	2020	2020

Following the ADVP curation criteria (see "**Data collection**", Figure 1), we first identified and screened 205 AD-related publications from 2009-2019. 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 (Section "**Meta-data design**") (converted into standard variant/association descriptors), annotated (Section "**Annotation**"), subjected to quality control steps (section "**Quality control steps**") and stored in ADVP (Figure 1).

ADVP data summary

The ADVP V1.0 release contains high-quality genome-wide and suggestive AD-related genetic associations extracted from GWAS publications. It contains 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.

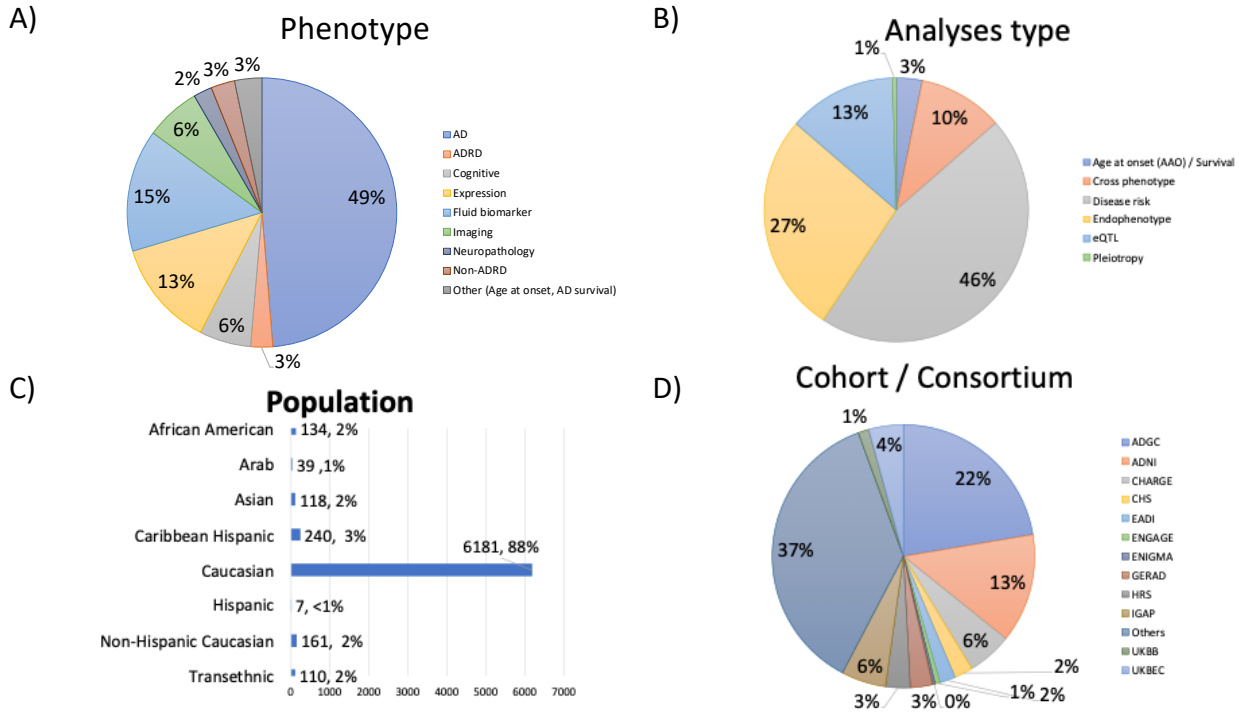


Figure 2: Summary of genetic association records in ADVP by A) Phenotype, B) Analyses type, C) Population, and D) Cohorts/Consortiums.

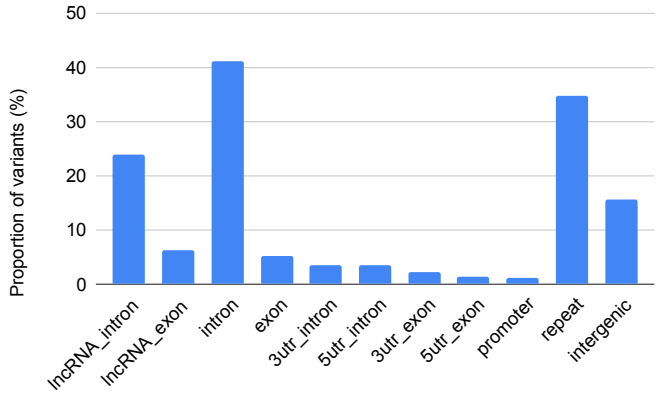
All ADVP association records are uniquely standardized into different categories:

1. As shown in Figure 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 (Figure 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 transethnic analyses. ~88% of the records are for Caucasian (Figure 2C).

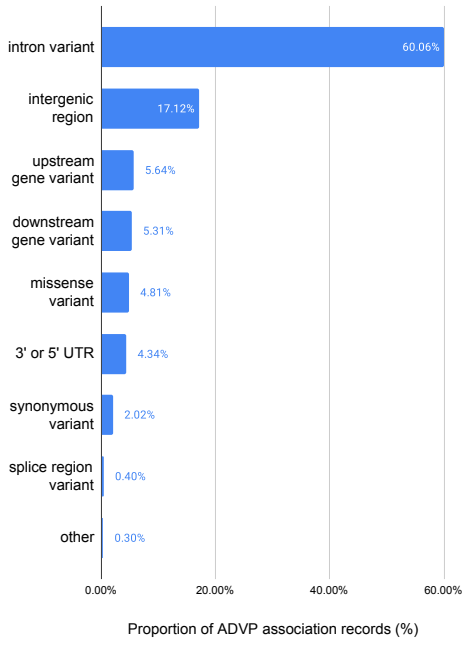
Others include African American, Arab, Asian, Caribbean Hispanic, Hispanic and Non-Hispanic Caucasian.

5. ADVP records span analyses results from over 80 cohorts. We summarize here data from the few largest AD consortiums (Figure 2D): 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) etc.

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 (based on computed normalization). ADVP associations are mostly located in non-coding regions including intronic (52.9%), intergenic (15.2%), and promoter (5.9%) (Figure 3A). ADVP records are also cross-referenced to NIAGADS Alzheimer's Genomics database (GenomicsDB, 2020). Figure 3B shows the impact of genetic variants in ADVP as determined by ADSP functional annotation pipeline (Butkiewicz et al., 2018; Cingolani et al., 2012; GenomicsDB, 2020).



A)



B)

Figure 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 (Butkiewicz et al., 2018; McLaren et al., 2016) and is provided by NIAGADS Alzhemer's Genomics database (GenomicsDB, 2020).

ADVP features – search, browse and visualize

ADVP aims to provide a simple and unified resource to the scientific community, allowing researches 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 described in **Methods**. Users can further select additional data fields via the column selector (Figure 4A). All records are integrated with the NIAGADS Alzheimer's Genomics database, allowing users to explore various kinds of biological annotations (e.g. CADD score (Rentzsch et al., 2019)) and functional genomics evidence, including overlaps with FANTOM5 (Andersson et al., 2014),

ENCODE histone modification (Dunham et al., 2012), and gene ontologies from KEGG (Kanehisa et al., 2016) and UniProt (Huntley et al., 2015).

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

1) By publication – users 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 ADVP records associated with these. Additionally, ADVP provides an interface for users to easily discover top variant or a gene with most association records or most publications via the summary counts for association records and papers (Figure 4B, 4C).
3) By region of interest – users can search and retrieve all genetic associations within the genomic regions of interest via this interface (https://advp.niagads.org/search).
4) By integrative genome-wide plots – investigators can navigate the landscape of AD genetics associations using the graphical display of genetic association data via interactive chromosome ideogram (https://advp.niagads.org/ideogram , Figure 4D) or interactive population/phenotype variant viewer (https://advp.niagads.org/plot , Figure 4E).

A)

Showing 1 - 12 of 12 records

Reset Filters | Show/Hide Columns

- Reported gene
- Interactions
- Population
- Population (detailed)
- Cohort
- Cohort (detailed)
- Sample size

Pubmed ID	Record type	SNP	Coordinates	Locus	Population	Cohort	Sample size	Subset Analyzed	Phenoty	
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	Disease risk
28714976	SNP-based	rs143332484	6:41129207	TREM2	Caucasian	CHARGE, EADI, GERAD	3,968	All	AD	Disease risk
28714976	SNP-based	rs75932628	6:41129252	TREM2	Caucasian	ADGC	14,884	All	AD	Disease risk
28714976	SNP-based	rs75932628	6:41129252	TREM2	Caucasian	ADGC, CHARGE, EADI, GERAD	30,018	All	AD	Disease risk
28714976	SNP-based	rs143332484	6:41129207	TREM2	Caucasian	ADGC, CHARGE, EADI, GERAD	33,786	All	AD	Disease risk
28714976	SNP-based	rs75932628	6:41129252	TREM2	Caucasian	CHARGE, EADI, GERAD	35,831	All	AD	Disease risk
28714976	SNP-based	rs143332484	6:41129207	TREM2	Caucasian	ADGC, CHARGE, EADI, GERAD	53,042	All	AD	Disease risk
28714976	SNP-based	rs75932628	6:41129252	TREM2	Caucasian	ADGC, CHARGE, EADI, GERAD	80,733	All	AD	Disease risk

« ‹ 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	View	39	16
rs11136000	CLU; CUGBP2; PICALM	chr8:27464519	View	60	15
rs2075650	APOC1, APOC4, APOE, PV	chr19:45395619	View	43	15
rs11218343	PICALM; SORL1	chr11:121435587	View	27	13
rs3851179	CUGBP2; PICALM	chr11:85868640	View	50	13
rs6656401	CR1,CR2; CR1; CR1L; CR2	chr1:207692049	View	76	12
rs3818361	CR1; CUGBP2	chr1:207784968	View	43	11
rs157580	PVRL2, TOMM40; TOMM4	chr19:45395266	View	15	10
rs1476679	PILRA, PILRB; GATS; MSX:	chr7:100004446	View	131	9
rs405509	APOE, TOMM40; APOE	chr19:45408836	View	28	9

« ‹ Page 1 of 183 › » Show 10 records

C)

Genes

Sort by frequency

Showing 1 - 10 of 952 records

Gene Name	Ensembl ID	View in Genomics DB	Association records	Papers
CLU	ENSG00000120885	View	179	29
BIN1	ENSG00000136717	View	138	28
CR1	ENSG00000203710	View	170	26
PICALM	ENSG00000073921	View	152	22
APOE	ENSG00000130203	View	99	22
TOMM40	ENSG00000130204	View	136	19
ABCA7	ENSG00000064687	View	82	18
SORL1	ENSG00000137642	View	164	17
EPHA1	ENSG00000146904	View	40	17
CD33	ENSG00000105383	View	40	16

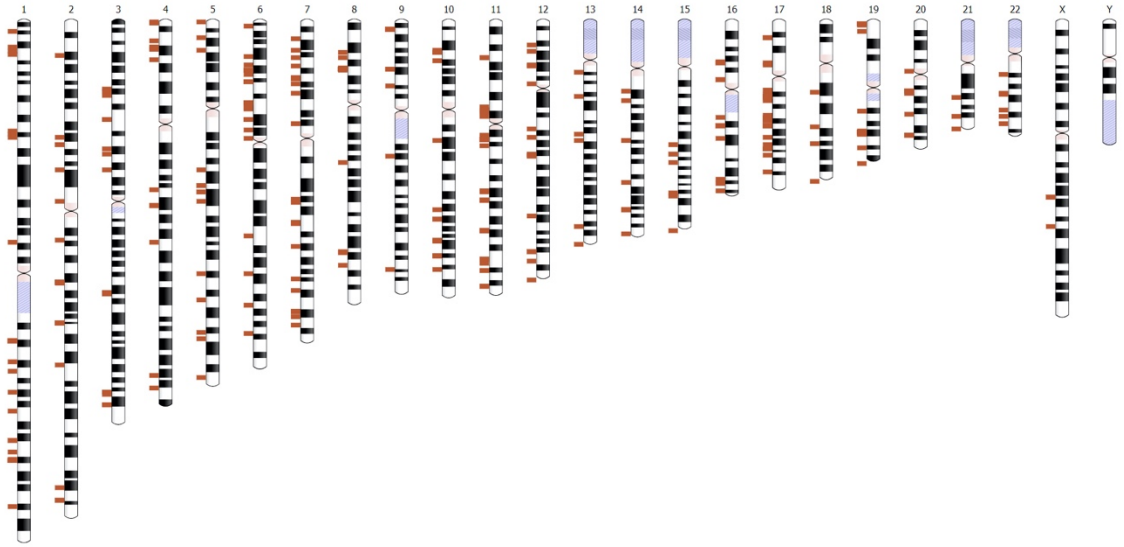
« ‹ Page 1 of 96 › »

Show 10 records

D)

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 chromosme view, click on the chromosome again, to go back to the ideogram.



Legend

ADVP genome-wide significant variants (p<5e-8)

E)

Variant viewer by Population/Phenotype

Switch between 'Population', 'Phenotype' to view variants across populations, and phenotypes.

To highlight variants for a particular population/phenotype, click on the corresponding population/phenotype in the legend (clicking on a legend item will toggle between highlight and normal view).

Population Phenotype

Select Effect size axis:

Log
Linear

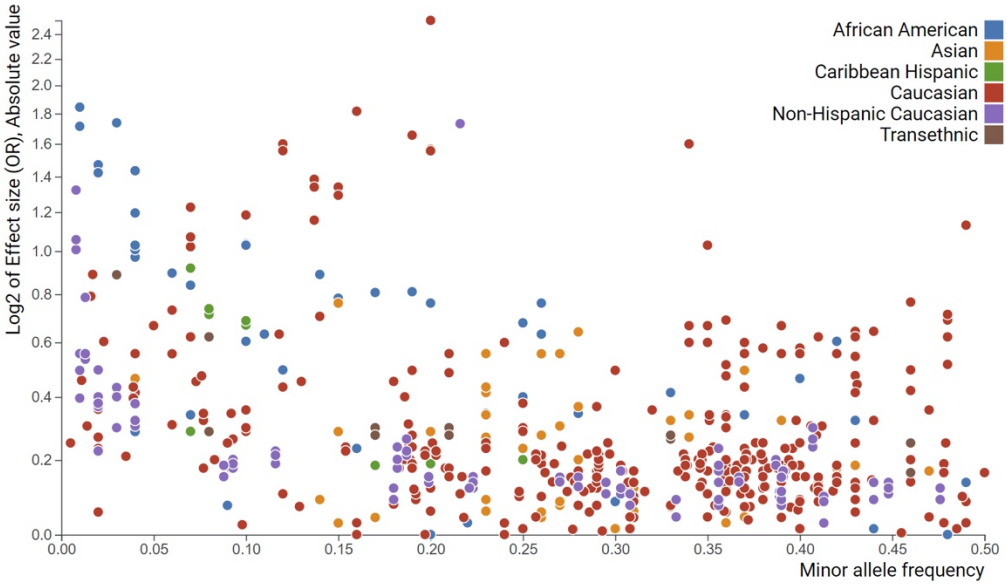


Figure 4 – ADVP interface. A) Association records table. The displayed information can be customized via column/field selector and filtered using provided text and data filters; B) Top variants curated in ADVP. Variants are displayed according to the number of reporting publications by default; Variant records and variant-related publications can be quickly accessed; C) Top genes curated in ADVP. Genes are ordered by the number of reporting publications by default; gene-related association records and publications can be quickly accessed; D) Interactive chromosome ideogram-based variant browser; E) 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.

ADVP use cases

ADVP is designed for various use cases in mind. First, ADVP can serve as an entry point for researchers starting 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, researchers 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, researchers can use ADVP to check if their findings have functional support from eQTL or other functional evidence collected in the NIAGADS Alzheimer's genomics database (Figure 5).

Figure 5 – Integration with NIAGADS AD Genomics database (GenomicsDB, 2020) providing additional biological information and functional evidence. Shown are the provided annotation and functional genomics data categories (red rectangles).

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 release, V1.0 (August 2020)). Among the main distinctive features of ADVP is the uniqueness of reporting harmonized AD variant and association information (standardized meta-table curation schema), integration with the genomic annotation, and functional information (NIAGADS Alzheimer's Genomics database (GenomicsDB, 2020)), as well as extensive consortium level information.

ADVP uniquely includes associations at SNP, gene, and interaction levels and contains curated phenotypes 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 (Figure 2B).

In addition to the standard *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, standardized gene names, study information (population, cohort, sample size, subset analyzed), and details of analyses (analyses type, imputation) (Figure 4A). All these columns enable the researchers to interpret, compare and view these records at different levels: phenotype (Figure 2A), population (Figure 2C), cohort (Figure 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 (Figure 3). They are also cross-referenced with NIAGADS Alzheimer's Genomics DB (GenomicsDB, 2020), providing other genomic annotation and functional genomic information. The standardized, structured design of ADVP association data allows systematic integration with other genetic, genomics, and molecular databases.

Lastly, we made substantial efforts to ensure high-quality of ADVP data contents. First, quality control at multiple levels is performed (Figure 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 versioned releases every six months. New publications on any AD-related GWAS studies will be added in an ongoing manner. It will first be added to our existing unified catalog ('Publications' meta-data table). Genetic association records will then be extracted from each of these publications, processed, QC-ed, and imported into ADVP (Association's meta-data table).

In the future, in addition to curating and including new AD GWAS findings, ADVP data collections will consist of a broader range of genetic results:

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

Last but not least, 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 researchers to easily access, interpret, compare, and visualize the vast collection of AD genetics findings.

Availability

All AD variant and association information is available through ADVP website (https://advp.niagads.org/). The code for processing reported variant and association data is also available upon request.

Acknowledgement

The authors thank Laura Cantwell, Lauren Kleidermacher, and Mitchell Tang for their help with this work. This work was supported by the National Institute on Aging [U24-AG041689, U54-AG052427, U01-AG032984]; 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.

Conflict of interest

The authors declare no conflict of interest.

References

Amlie-Wolf, A., Tang, M., Mlynarski, E. E., Kuksa, P. P., Valladares, O., Katanic, Z., Tsuang, D., Brown, C. D., Schellenberg, G. D., & Wang, L.-S. (2018). INFERNO: inferring the molecular mechanisms of noncoding genetic variants. *Nucleic Acids Research*, *46*(17), 8740–8753. https://doi.org/10.1093/nar/gky686

Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T., Ntini, E., Arner, E., Valen, E., Li, K.,

Schwarzfischer, L., Glatz, D., Raithel, J., Lilje, B., Rapin, N., … Sandelin, A. (2014). An atlas of active enhancers across human cell types and tissues. *Nature*, *507*(7493), 455–461. https://doi.org/10.1038/nature12787

Association, A. (2019). 2019 Alzheimer's disease facts and figures. *Alzheimer's & Dementia*, *15*(3), 321–387. https://doi.org/doi:10.1016/j.jalz.2019.01.010

Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., Korbel, J. O., Marchini, J. L., McCarthy, S., McVean, G. A., Abecasis, G. R., Auton, A., Abecasis, G. R., Altshuler, D. M., Durbin, R. M., Abecasis, G. R., Bentley, D. R., Chakravarti, A., Clark, A. G., Donnelly, P., … Abecasis, G. R. (2015). A global reference for human genetic variation. *Nature*, *526*(7571), 68–74.

Barral, S., Bird, T., Goate, A., Farlow, M. R., Diaz-Arrastia, R., Bennett, D. A., Graff-Radford, N., Boeve, B. F., Sweet, R. A., Stern, Y., Wilson, R. S., Foroud, T., Ott, J., Mayeux, R., Green, R., Kowall, N., Farrer, L., Williamson, J., Santana, V., … Anosike, E. (2012). Genotype patterns at PICALM, CR1, BIN1, CLU, and APOE genes are associated with episodic memory. *Neurology*, *78*(19), 1464–1471.

Barral, S., Cosentino, S., Christensen, K., Newman, A. B., Perls, T. T., Province, M. A., & Mayeux, R. (2014). Common genetic variants on 6q24 associated with exceptional episodic memory performance in the elderly. *JAMA Neurol*, *71*(12), 1514–1519.

Beecham, G. W., Hamilton, K., Naj, A. C., Martin, E. R., Huentelman, M., Myers, A. J., Corneveaux, J. J., Hardy, J., Vonsattel, J. P., Younkin, S. G., Bennett, D. A., De Jager, P. L., Larson, E. B., Crane, P. K., Kamboh, M. I., Kofler, J. K., Mash, D. C., Duque, L., Gilbert, J. R., … Montine, T. J. (2014). Genome-wide association meta-analysis of neuropathologic features of Alzheimer's disease and related dementias. *PLoS Genet.*, *10*(9), e1004606.

Bertram, L., McQueen, M. B., Mullin, K., Blacker, D., & Tanzi, R. E. (2007). Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database. *Nat. Genet.*, *39*(1), 17–23.

Biffi, A., Anderson, C. D., Desikan, R. S., Sabuncu, M., Cortellini, L., Schmansky, N., Salat, D., & Rosand, J. (2010). Genetic variation and neuroimaging measures in Alzheimer disease. *Arch. Neurol.*, *67*(6), 677–685.

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 Research*, *47*(D1), D786--D792. https://doi.org/10.1093/nar/gky930

Buniello, A., MacArthur, J. A. L., Cerezo, M., Harris, L. W., Hayhurst, J., Malangone, C., McMahon, A., Morales, J., Mountjoy, E., Sollis, E., Suveges, D., Vrousgou, O., Whetzel, P. L., Amode, R., Guillen, J. A., Riat, H. S., Trevanion, S. J., Hall, P., Junkins, H., … Parkinson, H. (2019). The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.*, *47*(D1), D1005–D1012.

Butkiewicz, M., Blue, E. E., Leung, Y. Y., Jian, X., Marcora, E., Renton, A. E., Kuzma, A., Wang, L. S., Koboldt, D. C., Haines, J. L., & Bush, W. S. (2018). Functional annotation of genomic variants in studies of late-onset Alzheimer's disease. *Bioinformatics*, *34*(16), 2724–2731. https://doi.org/10.1093/bioinformatics/bty177

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., & Ruden, D. M. (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*, *6*(2), 80–92. https://doi.org/10.4161/fly.19695

Corder, E. H., Saunders, A. M., Strittmatter, W. J., Schmechel, D. E., Gaskell, P. C., Small, G. W., Roses, A. D., Haines, J. L., & Pericak-Vance, M. A. (1993). Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science (New York, N.Y.)*, *261*(5123), 921–923. https://doi.org/10.1126/science.8346443

Cruchaga, C., Kauwe, J. S., Harari, O., Jin, S. C., Cai, Y., Karch, C. M., Benitez, B. A., Jeng, A. T., Skorupa, T., Carrell, D., Bertelsen, S., Bailey, M., McKean, D.,

Shulman, J. M., De Jager, P. L., Chibnik, L., Bennett, D. A., Arnold, S. E., Harold, D., … Drost, D. (2013). GWAS of cerebrospinal fluid tau levels identifies risk variants for Alzheimer's disease. *Neuron*, *78*(2), 256–268.

Cukier, H. N., Kunkle, B. W., Vardarajan, B. N., Rolati, S., Hamilton-Nelson, K. L., Kohli, M. A., Whitehead, P. L., Dombroski, B. A., Van Booven, D., Lang, R., Dykxhoorn, D. M., Farrer, L. A., Cuccaro, M. L., Vance, J. M., Gilbert, J. R., Beecham, G. W., Martin, E. R., Carney, R. M., Mayeux, R., … Pericak-Vance, M. A. (2016). ABCA7 frameshift deletion associated with Alzheimer disease in African Americans. *Neurol Genet*, *2*(3), e79.

Dunham, I., Kundaje, A., Aldred, S. F., Collins, P. J., Davis, C. A., Doyle, F., Epstein, C. B., Frietze, S., Harrow, J., Kaul, R., Khatun, J., Lajoie, B. R., Landt, S. G., Lee, B. K., Pauli, F., Rosenbloom, K. R., Sabo, P., Safi, A., Sanyal, A., … Lochovsky, L. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*, *489*(7414), 57–74. https://doi.org/10.1038/nature11247

Frankish, A., Diekhans, M., Ferreira, A. M., Johnson, R., Jungreis, I., Loveland, J., Mudge, J. M., 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, I. T., … Flicek, P. (2019). GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res.*, *47*(D1), D766--D773.

Gatz, M., Pedersen, N. L., Berg, S., Johansson, B., Johansson, K., Mortimer, J. A., Posner, S. F., Viitanen, M., Winblad, B., & Ahlbom, A. (1997). Heritability for Alzheimer's disease: The study of dementia in Swedish twins. *Journals of Gerontology - Series A Biological Sciences and Medical Sciences*, *52*(2). https://doi.org/10.1093/gerona/52A.2.M117

Genin, E., Hannequin, D., Wallon, D., Sleegers, K., Hiltunen, M., Combarros, O., Bullido, M. J., Engelborghs, S., De Deyn, P., Berr, C., Pasquier, F., Dubois, B., Tognoni, G., Fi?vet, N., Brouwers, N., Bettens, K., Arosio, B., Coto, E., Del Zompo, M., … Campion, D. (2011). APOE and Alzheimer disease: a major gene with semi-

dominant inheritance. *Mol. Psychiatry*, *16*(9), 903–907.

GenomicsDB. (2020). *NIAGADS Alzheimer's Genomics database*. http://www.niagads.org/genomics/

Harold, D., Abraham, R., Hollingworth, P., Sims, R., Gerrish, A., Hamshere, M. L., Pahwa, J. S., Moskvina, V., Dowzell, K., Williams, A., Jones, N., Thomas, C., Stretton, A., Morgan, A. R., Lovestone, S., Powell, J., Proitsi, P., Lupton, M. K., Brayne, C., … Williams, J. (2009). Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nat. Genet.*, *41*(10), 1088–1093.

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., … Kanba, S. (2015). A genome-wide association study of late-onset Alzheimer's disease in a Japanese population. *Psychiatr. Genet.*, *25*(4), 139–146.

Hollingworth, P., Harold, D., Sims, R., Gerrish, A., Lambert, J. C., Carrasquillo, M. M., Abraham, R., Hamshere, M. L., Pahwa, J. S., Moskvina, V., Dowzell, K., Jones, N., Stretton, A., Thomas, C., Richards, A., Ivanov, D., Widdowson, C., Chapman, J., Lovestone, S., … Williams, J. (2011). Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. *Nat. Genet.*, *43*(5), 429–435.

Huntley, R. P., Sawford, T., Mutowo-Meullenet, P., Shypitsyna, A., Bonilla, C., Martin, M. J., & O'Donovan, C. (2015). The GOA database: Gene Ontology annotation updates for 2015. *Nucleic Acids Research*, *43*(D1), D1057--D1063. https://doi.org/10.1093/nar/gku1113

Jansen, I. E., Savage, J. E., Watanabe, K., Bryois, J., Williams, D. M., Steinberg, S., Sealock, J., Karlsson, I. K., H?gg, S., Athanasiu, L., Voyle, N., Proitsi, P., Witoelar, A., Stringer, S., Aarsland, D., Almdahl, I. S., Andersen, F., Bergh, S., Bettella, F., … Posthuma, D. (2019). Genome-wide meta-analysis identifies new loci and

functional pathways influencing Alzheimer's disease risk. *Nat. Genet.*, *51*(3), 404–413.

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., & Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Research*, *44*(D1), D457--D462. https://doi.org/10.1093/nar/gkv1070

Kauwe, J. S., Bailey, M. H., Ridge, P. G., Perry, R., Wadsworth, M. E., Hoyt, K. L., Staley, L. A., Karch, C. M., Harari, O., Cruchaga, C., Ainscough, B. J., Bales, K., Pickering, E. H., Bertelsen, S., Fagan, A. M., Holtzman, D. M., Morris, J. C., & Goate, A. M. (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*(10), e1004758.

Kuksa, P. P., Lee, C.-Y., Amlie-Wolf, A., Gangadharan, P., Mlynarski, E. E., Chou, Y.-F., Lin, H.-J., Issen, H., Greenfest-Allen, E., Valladares, O., Leung, Y. Y., & Wang, L.-S. (2020). SparkINFERNO: A scalable high-throughput pipeline for inferring molecular mechanisms of non-coding genetic variants. *Bioinformatics*. https://doi.org/10.1093/bioinformatics/btaa246

Kunkle, B. W., Grenier-Boley, B., Sims, R., Bis, J. C., Damotte, V., Naj, A. C., Boland, A., Vronskaya, M., van der Lee, S. J., Amlie-Wolf, A., Bellenguez, C., Frizatti, A., Chouraki, V., Martin, E. R., Sleegers, K., Badarinarayan, N., Jakobsdottir, J., Hamilton-Nelson, K. L., Moreno-Grau, S., … Pericak-Vance, M. A. (2019). Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Ab, tau, immunity and lipid processing. *Nat. Genet.*, *51*(3), 414–430.

Lambert, J. C., Heath, S., Even, G., Campion, D., Sleegers, K., Hiltunen, M., Combarros, O., Zelenika, D., Bullido, M. J., Tavernier, B., Letenneur, L., Bettens, K., Berr, C., Pasquier, F., Fi?vet, N., Barberger-Gateau, P., Engelborghs, S., De Deyn, P., Mateo, I., … Pilotto, A. (2009). Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. *Nat. Genet.*, *41*(10), 1094–1099.

Lambert, J. C., Ibrahim-Verbaas, C. A., Harold, D., Naj, A. C., Sims, R., Bellenguez, C., DeStafano, A. L., Bis, J. C., Beecham, G. W., Grenier-Boley, B., Russo, G., Thorton-Wells, T. A., Jones, N., Smith, A. V, Chouraki, V., Thomas, C., Ikram, M. A., Zelenika, D., Vardarajan, B. N., … Pickering-Brown, S. (2013). Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat. Genet.*, *45*(12), 1452–1458.

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., … Moore, H. F. (2013). The Genotype-Tissue Expression (GTEx) project. In *Nature Genetics* (Vol. 45, Issue 6, pp. 580–585). Nat Genet. https://doi.org/10.1038/ng.2653

McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R. S., Thormann, A., Flicek, P., & Cunningham, F. (2016). The Ensembl Variant Effect Predictor. *Genome Biology*, *17*(1). https://doi.org/10.1186/s13059-016-0974-4

Mez, J., Chung, J., Jun, G., Kriegel, J., Bourlas, A. P., Sherva, R., Logue, M. W., Barnes, L. L., Bennett, D. A., Buxbaum, J. D., Byrd, G. S., Crane, P. K., Ertekin-Taner, N., Evans, D., Fallin, M. D., Foroud, T., Goate, A., Graff-Radford, N. R., Hall, K. S., … Farrer, L. A. (2017). Two novel loci, COBL and SLC10A2, for Alzheimer's disease in African Americans. *Alzheimers Dement*, *13*(2), 119–129.

Naj, A. C., Jun, G., Beecham, G. W., Wang, L. S., Vardarajan, B. N., Buros, J., Gallins, P. J., Buxbaum, J. D., Jarvik, G. P., Crane, P. K., Larson, E. B., Bird, T. D., Boeve, B. F., Graff-Radford, N. R., De Jager, P. L., Evans, D., Schneider, J. A., Carrasquillo, M. M., Ertekin-Taner, N., … Schellenberg, G. D. (2011). Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. *Nat. Genet.*, *43*(5), 436–441.

Naj, A. C., & Schellenberg, G. D. (2017). Genomic variants, genes, and pathways of Alzheimer's disease: An overview. *Am. J. Med. Genet. B Neuropsychiatr. Genet.*, *174*(1), 5–26.

Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J., & Kircher, M. (2019). CADD: Predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Research*, *47*(D1), D886--D894. https://doi.org/10.1093/nar/gky1016

Seshadri, S., Fitzpatrick, A. L., Ikram, M. A., DeStefano, A. L., Gudnason, V., Boada, M., Bis, J. C., Smith, A. V, Carassquillo, M. M., Lambert, J. C., Harold, D., Schrijvers, E. M., Ramirez-Lorca, R., Debette, S., Longstreth, W. T., Janssens, A. C., Pankratz, V. S., Dartigues, J. F., Hollingworth, P., … Furberg, C. (2010). Genome-wide analysis of genetic loci associated with Alzheimer disease. *JAMA*, *303*(18), 1832–1840.

Sherry, S. T., Ward, M. H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M., & Sirotkin, K. (2001). dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.*, *29*(1), 308–311.

Widenius, M., Axmark, D., & DuBois, P. (2002). *Mysql Reference Manual* (1st ed.). O'Reilly & Associates, Inc.

Yates, A. D., Achuthan, P., Akanni, W., Allen, J., Allen, J., Alvarez-Jarreta, J., Amode, M. R., Armean, I. M., Azov, A. G., Bennett, R., Bhai, J., Billis, K., Boddu, S., Marugan, J. C., Cummins, C., Davidson, C., Dodiya, K., Fatima, R., Gall, A., … Flicek, P. (2020). Ensembl 2020. *Nucleic Acids Res.*, *48*(D1), D682--D688.

Supplementary information

Supplementary Table S1 – Curated publications and data sources
Supplementary Table S2 – All curated and harmonized/derived data fields in ADVP
Supplementary Table S3 – ADGC cohorts included in ADVP
Supplementary Figure S1 – Classification of studies by the analysis type and stage of analyses