Supplementary Material

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

Supplementary Note 1

Details on the nine meta-information data fields describing each association:

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 3), 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 1). 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").

Supplementary Note 2

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 standardized meta-data schema (see Meta-data curation in the main text). The ADVP web server runs on Amazon Web Services (AWS) cloud computing instance (m5.4xlarge) using MySQL [1] 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 (Fig. 1 in the main text) including publications, variant, and association tables. The web front-end provides multiple data views for publications, genes, variants, and association records (Fig. 4 in the main text).

REFERENCE

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

Supplementary Table 1. Curated publications and data sources. See separate Excel file.

Supplementary Table 2. All curated and harmonized/derived data fields in ADVP. See separate Excel file.

Supplementary Table 3. ADGC cohorts included in ADVP. See separate Excel file.

Supplementary Table 4. Top AD-associated loci and SNPs across populations. See separate Excel file.

Supplementary Figure 1. Classification of studies by the analysis type and stage of analyses.