**S1 | Supplemental methods**

**Study samples**

**Discovery samples:**

*NIH LOAD and NCRAD samples*

Subjects were obtained from NCRAD, and ascertainment and collection details can be found at the NCRAD website (http://www.ncrad.org). For our study, we used subjects of self-reported European ancestry with DNA selected from one affected individual from each family. Affected individuals were selected based on strength of diagnosis (confirmed and probable) and DNA availability. For both cohorts (NIH LOAD or NCRAD), cases and families were required to meet the same diagnostic study criteria. In the families, probands were required to have a diagnosis of definite or probable LOAD with onset > 60 years of age. Both studies also required additional family members with definite, probable or possible AD. The NIH LOAD study also includes unrelated controls without family history of AD that were required to have cognitive testing and clinical examination to establish normal cognitive functioning for age and absence of AD.

*NACC samples*

The NACC was established by the National Institute on Aging to facilitate collaborative research. Samples were collected from 27 NIH-funded Alzheimer's Disease Centers (ADCs) across the United States. Cases and controls describing themselves as “white” that have met stringent gold-standard diagnostic criteria as described before 1–3 were selected for targeted sequencing. The NACC database September 2013 freeze was used to select cases and controls. The DNA samples were obtained from the National Cell Repository for Alzheimer's Disease (NCRAD).

*NIMH cases*
NIMH families have been obtained by three sites (University of Alabama - Birmingham, Johns Hopkins University, and Massachusetts General Hospital/Harvard Medical School). Data collection has been coordinated among the three sites by using a common protocol that includes uniform assessments and medical, neurologic, and psychiatric histories. Operational criteria for the clinical diagnosis of probable or possible AD following NINCDS-ADRDA Work Group guidelines have been implemented by all three sites. Each family had at least two affected individuals. Definite AD according to age-adjusted Khachaturian criteria was established on autopsy \(^4,5\). From each family one subject was selected for sequencing.

**NIMH controls**

In order to enable case-control association studies the NIMH has established a set of controls. This control sample consists of a group of adults who completed an online, short self-report clinical assessment. For our study we included Caucasian individuals older than 60 years. NIMH controls did not have cognitive testing \(^6\).

**ACT controls**

The Adult Changes in Thought (ACT) study enrolled cognitively intact, age 65 and older, members of the Group Health Cooperative (GHC) health care system in the Seattle, Washington, USA \(^7\). All of the subjects included in our study were cognitively normal at the age of the last exam.

**Washington University**

The samples were obtained at the Knight-ADRC at Washington University. Cases and controls were evaluated in the Clinical Core of the Knight ADRC. Cases received a clinical diagnosis of AD dementia in accordance with standard criteria \(^3\) and dementia severity was determined with the Clinical Dementia Rating (CDR) \(^8\). All individuals were of European descent.
In all cohorts written consent was obtained from all participants at participating institutions.

**Replication study:**

**German Sample**

The German AD cases for genotyping were recruited from the following three sources: (i) the German Dementia Competence Network (DCN, n=311); (ii) the German study on Aging, Cognition, and Dementia in primary care patients (AgeCoDe, n=148); and (iii) the interdisciplinary Memory Clinic at the University Hospital of Bonn (n=692). The German controls were drawn from the AgeCoDe cohort (n=851). (i) The DCN cases were evaluated using extensive neuropsychological tests including those of the Consortium to Establish a Registry for Alzheimer’s Disease (CERAD); the Mini Mental State Exam (MMSE); and the Clinical Dementia rating (CDR-SB). Whole blood, genomic DNA, plasma, and CSF were collected from all participants. (ii) AgeCoDe is a German multicenter prospective cohort study. The assessment includes the Structured Interview for Diagnosis of Dementia of Alzheimer type, Multi-infarct Dementia, and Dementia of other etiology according to DSM-IV and ICD-10 (SIDAM). Dementia was diagnosed according to DSM-IV criteria and for AgeCoDe controls included in the present study, dementia and mild cognitive impairment had been excluded at the last follow-up visit. (iii) In interdisciplinary Memory Clinic of the Department of Psychiatry and Department of Neurology at the University Hospital in Bonn the diagnoses were assigned according the NINCDS/ADRDA criteria and on the basis of clinical history, physical examination, neuropsychological testing (using the CERAD neuropsychological battery, including the MMSE), laboratory assessments, and brain imaging.

**Finnish Sample**

The Finnish-AD cohort comprises of 672 AD patients and 686 age-matched healthy controls from Kuopio in Finland. All patients were diagnosed with probable AD according to the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer’s Disease and Related
Disorders Association (NINCDS-ADRDA) criteria. Control subjects had no symptoms of cognitive impairment based on clinical interview and neuropsychological examination.

**ADGC sample**

Cases and controls were taken from multiple ADGC datasets, described here briefly. The NACC National Institute on Aging sample, the University of Toronto/GlaxoSmithKline case-control sample, the Vanderbilt/Miami/Mount Sinai case-control sample, the National Institute on Aging LOAD sample, the NCRAD multiplex family-based sample, the Multi-institutional Research in Alzheimer's Genetic Epidemiology family-based sample, and ACT prospective cohort were described in references. The Genetics Differences cohort is a population-based prevalent case-control study from the same population as the ACT study. Additional studies included in ADGC sample are; The Washington Heights–Inwood Columbia Aging Project, Northern European ancestry subjects; the Washington University cohort is a case-control cohort of individuals with Northern European ancestry; The University of Miami case-control dataset; and the Cache County Study on Memory Health and Aging. Per individual source studies, all subjects were recruited under protocols approved by the appropriate Institutional Review Boards. Cases living at time of recruitment were adjudicated as possible or probable AD prior to analyses according to NINCDS/ADRDA criteria whereas affection status of all deceased cases was confirmed through autopsy.

**MIP design, sequencing and alignment**

MIPs were designed with the MIP generator pipeline. The targeted sequence for each MIP was 112 bp long with 20 bp extension and ligation arms. Each MIP included a 5 bp long sequence tag as an identification barcode to separate up to 1024 unique reads from PCR duplicates. Captured sequences were PCR amplified with primers containing an 8 bp barcode to distinguish individual samples. MIPs were at first combined in equimolar amounts. Target capture and sequencing was performed as described. A subset of samples were sequenced on Illumina MiSeq® to allow recalibration of probes depending upon probe performance. The recalibrated MIPs were divided in two pools based
on performance. Capturing and barcoding was performed for each sample in both pools and barcoded were merged with in 1:3 ratio.

192 samples, divided between cases and controls were sequenced one lane of an Illumina Hiseq® 2500 sequencer. Resulting reads were checked for quality with FastQC. The split_libraries_fastq.py script from the tool kit Qiime was used to identify the barcodes in each read and after identification the reads were written back into a FASTQ file.

Reads were aligned to the human genome (hg19) with BWA. The aligned file was divided into 192 unique barcoded samples and separated into FASTQ files. Prinseq was used to split extracted reads into unique and duplicates followed by duplicate and MIP arm removal. Resulting FASTQ files were again aligned to hg19 and individual samples separated. Individual alignments were further refined using the GATK ‘IndelRealigner’ and ‘BaseRecalibrator’ tools.

**Variant calling and genotyping**

Next generation sequencing coverage was determined with the GATK DepthOfCoverage tool. Variants were called and filtered (QualByDepth < 1.5, HomopolymerRun > 6 and Total Depth over all samples ≥ 20) with the Genome Analysis Toolkit (GATK) [28]. Individuals that could not be genotyped for ≥ 95% of the variants and variants that could not be found in ≥ 95% of the remaining individuals were removed before analysis. To exclude possible sequencing artifacts, we followed the GATK best practices workflow for variant discovery. Variants that passed QC were annotated for frequency and function with the SeattleSeq 137 annotation pipeline. Variant frequency was assigned based on 1000 Genome European dataset (phase1_release_v3.20101123). All annotated variants had at least 8 reads and at least 25% of reads (minimum two) had to show the alternate allele to be called heterozygous.

In the Finish and German samples genotyping was performed using Sequenom iPLEXa and TaqMan® SNP genotyping assay respectively. The variants in the CASP8 and APH1B genes in cases and controls were confirmed with capillary sequencing in the primary study and in the German replication sample.
For ADGC samples 18,099 samples were genotyped on exome arrays, including 16,571 samples on the Infinium HumanExome V1 Beadchip (Illumina, Inc) and 1,528 samples on the V3 Beadchip. Genotyping was performed for 8,410 samples at the Robert S. Boas Center for Genomics and Human Genetics, Feinstein Institute for Medical Research, Manhasset, New York; 1,990 individuals at the John P. Hussman Institute for Human Genomics, University of Miami, Miami, Florida; and 7,694 individuals at the Center for Applied Genomics, The Children’s Hospital of Philadelphia, Philadelphia, Pennsylvania. Genotypes were initially called using the default clustering profiles from Illumina and recalled using clustering profiles generated by Genentech using data from 30,000 samples.

Exome chip genotyping data (18,099 samples total) were first pre-processed using quality check steps adapted from Naj et al. Briefly, 2,275 samples were excluded due to: (a) overall genotype call rates <98%, (b) genotype-imputed or reported sex mismatch, (c) excess relatedness (\(\pi^e > 0.4\) using 15,086 linkage disequilibrium–pruned autosomal markers with a minor allele frequency [MAF]>0.1), or (d) lack of Northern European ancestry. Non-Caucasian samples were identified as outliers during population substructure estimation using the program EIGENSTRAT for principal components analysis. In all, 8,390 AD cases and 7,434 cognitively-normal individuals passed quality control and used in analysis in our study.


