Genomics of Alzheimer's disease: Value of high-throughput genomic technologies to dissect its etiology

Giuseppe Tosto^{1,2,3}, Christiane Reitz^{1,2,3,4*}

¹ The Taub Institute for Research on Alzheimer's Disease and the Aging Brain, Columbia University, New York, New York, USA

² The Gertrude H. Sergievsky Center, Columbia University, New York, New York, USA

³ The Department of Neurology, Columbia University, New York, New York, USA

⁴ The Dept. of Epidemiology, Columbia University, New York, New York, USA

WORD COUNT: 3,507

RUNNING HEADER: Genomics of Alzheimer's disease

KEY WORDS: Genetics; Genome-wide association study; Alzheimer's disease; high-throughput technologies; rare variants; epidemiology of neurodegenerative disorders.

*Correspondence

ABSTRACT

Late-onset Alzheimer's disease (AD), the most common neurodegenerative disorder in western countries, is clinically defined by progressive worsening in cognitive functions along with function and behavioral impairment. This ultimately results in complete incapacity and death. AD is a clinically and pathologically heterogeneous disease, and this is reflected by the numerous genetic findings that point to several diverse molecular mechanisms and pathways. Linkage, genome-wide association and next-generation sequencing studies have led to the identification of more than 20 novel susceptibility loci for AD. While these observations have significantly increased the knowledge of pathogenic mechanisms and potential therapeutic targets, a large part of the genetic component underlying AD is still unexplained. This review will summarize and discuss the impact of these advances on AD diagnosis and prediction of prognosis.

INTRODUCTION

In the US alone, about 5.5 million individuals are affected with late-onset Alzheimer's disease (AD)¹ and this number is destined to increase dramatically in the next few decades. The predominant clinical features of AD are a progressive, and ultimately fatal, worsening in cognition, function and behavior, but a broad phenotypic profile is commonly observed both in terms of timing of onset and rapidity of progression². Pathological findings at brain autopsy are intracellular deposits of hyper-phosphorylated tau protein in the form of neurofibrillary tangles, and extracellular β -amyloid (A β) protein in diffusible oligomers and insoluble plaques. Neuropathological findings are heterogeneous as well, with common observation of Lewy bodies and/or vascular changes². Available drugs mostly belong to two classes of compounds, namely acetylcholinesterase inhibitors and NMDA receptor antagonists. Both are widely used and are relatively safe to use, but systematic reviews (i.e. Cochrane reviews) failed to demonstrate any significant benefit in terms of disease prevention or in slowing of disease progression. Mapping the genes and molecular pathways linked to the disease will be critical for identifying more effective preventive and therapeutic targets.

IDENTIFICATION OF *APP*, *PSEN1* AND *PSEN2* AS AD RISK GENES BY EARLY LINKAGE STUDIES

Autosomal dominant mutations in the *APP*, *PSEN1* and *PSEN2* genes,³⁻⁵ were discovered by linkage studies conducted in large families multiply affected by early-onset AD (onset age: <50 years old) in the 1990's. A genetic linkage study is a family-based method used to map a disease or trait to a genomic location by demonstrating its co-segregation with genetic markers

throughout a pedigree. This approach leverages the fact that co-segregating genomic regions are more likely to contain a causal genetic variant compared to regions without co-segregation.

The successful mapping of *APP*, *PSEN1* and *PSEN2* genes as disease loci helped to formulate the ‘amyloid cascade hypothesis’. The latter indicated as a key pathogenic step the enhanced generation of Amyloid beta protein (A β) fragments generated through alternative pathogenic cleavage of the amyloid precursor protein (APP) by β - and γ -secretases, as opposed to the non-amyloidogenic cleavage by α -secretase. A β peptides generated through the amyloidogenic pathway are prone to aggregation, leading to toxic A β oligomers and amyloid plaque formation. As of to date (March 2016), 49 pathogenic mutations in *APP*, 216 mutations in *PSEN1* and 16 mutations in *PSEN2* have been mapped (<http://www.molgen.ua.ac.be/ADMutations/>).

Linkage studies in multiplex AD families also led to identification of the first susceptibility gene identified in the common, late-onset form of AD (LOAD), i.e. Apolipoprotein E (*APOE*) in 1993.^{6,7} In humans, the *APOE* gene has 3 exons and is located on chromosome 19; it has three isoforms exhibiting major differences in effects on structure and function of *APOE*. These isoforms are encoded by the *APOE* ϵ 2, ϵ 3, and ϵ 4 alleles.

A single *APOE* ϵ 4 allele increases LOAD risk by ~3 fold^{8,9} and lowers memory function as well as the age at onset of AD.^{6,10-13} The mechanisms through which *APOE* increases AD risk is not fully clarified. In peripheral tissues, *APOE* is primarily produced by the liver and macrophages, and mediates blood lipids and cardiovascular risk profile in an isoform-dependent manner.

Bennet et al. showed that *APOE*- ϵ 2 is associated with lower LDL and reduced risk of coronary artery disease, whereas the presence of the ϵ 4 allele increased both.¹⁴ In the central nervous system, *APOE* traffics cholesterol resulting from neurodegeneration to neurons requiring them

for membrane repair, proliferation, or remyelination. In addition, APOE affects glutamate receptor function and synaptic plasticity by modulating neuronal APOE receptor recycling¹⁵. *APOE-ε4* dosage at autopsy correlates strongly with Aβ load and plaque accumulation^{16,17}. Correspondingly, in *in vitro* studies the *APOE-ε4* isoform binds Aβ peptides more vigorously than *APOE-ε3*¹⁸ and leads to increased Aβ aggregation^{19,20}; knockout *APOE*^{−/−} mice develop less nonfibrillar Aβ deposits^{21,22}. APOE may also indirectly affect clearance of Abeta by competing for the same astrocyte receptor, low-density lipoprotein receptor-related protein 1 (LRP1), which helps clear Aβ,²³ Finally, APOE4 may stimulate tau phosphorylation leading to pre-neurofibrillary tangles²⁴. Notably, there are important differences among ethnic groups: APOE-ε4 association with AD is in fact weaker among African Americans and Hispanics²⁵. Our group reported that Caribbean Hispanics have an increased prevalence of LOAD regardless of their APOE genotype; in the absence of an ε4 allele, African American and Hispanics show 2 to 4 times fold higher incidence of LOAD as compared to the Whites.²⁵ Moreover, we also showed that, among Caribbean Hispanics, familial LOAD is strongly associated with APOE-ε4 whereas sporadic LOAD has only weak association with the risk allele.²⁶

SUSCEPTIBILITY LOCI ASSOCIATED WITH LATE-ONSET AD IDENTIFIED BY HIGH-THROUGHPUT TECHNOLOGIES

Common disorders such as the late-onset form of AD are caused by a combination of multiple environmental risk factors and common and rare genetic variants in genes belonging to various pathogenic pathways (“polygenic disease”, “complex disorder”). Under this scenario, a genetic association study can demonstrate higher statistical power than the linkage approach if the sample size is sufficiently large.

Over the past decade, the development of high-throughput massively parallel technologies for disentangling disease at the molecular level has revolutionized medical science and has significantly advanced the understanding of the molecular underpinnings of AD. These “next generation” genomic technologies include genome-wide genotyping arrays for genome-wide association studies (GWAS), whole-exome (WES) or whole-genome (WGS) sequencing for detection of disease-associated DNA sequence variants (“genomics”), and RNA sequencing (RNA-seq) for transcriptome and non-coding RNA analysis (“transcriptomics”).

AD susceptibility loci identified by GWAS studies. By definition, genotyping chips for GWAS are designed to examine common variation with a minor allele frequency (MAF) > ~1-5% across the whole genome. The large AD GWAS performed to date focused mostly on individuals of non-Hispanics Whites of European ancestry²⁷⁻³¹. In addition to *APOE*, over 20 novel loci have been identified by these studies. Common to all these loci are the associated small effect sizes with Odd Ratios between 1.1 and 1.3, suggesting that a large part of the genetic component of LOAD is still unexplained. The largest GWAS to date, conducted in 2013 by the International Genomics of Alzheimer's Project (IGAP) aggregated 74,046 non-Hispanic Whites³² through a multi-center collaboration. The study confirmed *CRI*, *BIN1*, *CD2AP*, *EPHA1*, *CLU*, *MS4A6A*, *PICALM*, *ABCA7* and *CD33*, previously identified by the first set of GWAS, at genome-wide significance and identified in addition *HLA-DRB5/HLA-DRB1*, *PTK2B*, *SORL1*, *SLC24A4/RIN3*, *DSG2*, *INPP5D*, *MEF2C*, *NME8*, *ZCWPW1*, *CELF1*, *FERMT2* and *CASS4*.³³ Most of these genes cluster in a specific set of pathways, namely immune response (*HLA-DRB5/DRB1*, *INPP5D*, *MEF2C*), APP processing (*SORL1*, *CASS4*), Tau pathology (*CASS4*, *FERMT2*), cell migration (*PTK2B*) and lipid transport and endocytosis (*SORL1*) strongly reinforcing the importance of these molecular mechanisms in AD etiology. Of note, *SORL1* (sortilin-related

receptor, L (DLR class) 1) had previously been demonstrated to modulate trafficking and processing of APP in a candidate gene approach³³. In addition, consistent with the notion of a complex disease, the findings of this multicenter project further strengthen the evidence for additional molecular mechanisms including hippocampal synaptic function (*MEF2C*, *PTK2B*); cytoskeletal function and axonal transport (*CELF1*, *NME8*, *CASS4*); regulation of gene expression and post-translational modification of proteins, microglial and myeloid cell function (*INPPD5*), and APP (*SORL1* and *CASS4*) and tau (*CASS4* and *FERMT2*) pathology.

Due to differences in allele frequencies and LD patterns, loci detected by genomic studies can differ between ethnic groups. In the largest GWAS performed to date in African Americans, *ABCA7* was confirmed as an AD susceptibility locus, notably with an effect size similar to that of *APOE*³⁴. In addition, *CRI*, *BIN1*, *EPHA1* and *CD33* were confirmed in gene-based analyses. We recently published a large GWAS in Caribbean Hispanics where a novel locus in the *FBXL7* gene was found, along with other known-loci previously identified in large GWAS of European ancestry³⁵. The *FBXL7* gene encodes a subcellular protein involved in phosphorylation-dependent ubiquitination processes and displays pro-apoptotic activity. F-box proteins also modulate inflammation and innate immunity, pathways already implicated in LOAD by several of the above listed genes. This new finding implicates additional mechanisms underlying the pathophysiology of LOAD and demonstrates the valuable asset of admixed populations in the understanding of the physiopathology of the disease. A study conducted in Han Chinese also confirmed *PICALM*, *BIN1*, *CLU*, and *MS4A4E* as susceptibility loci for AD³⁶.

Beside single-marker analyses, haplotype association analyses have identified additional susceptibility loci. Lambert and colleagues³⁷ conducted a sliding-windows genome-wide search

which led to the identification of a haplotype within the *FRMD4A* gene (Chr.10p13).

Furthermore, SNPs lying within the *FRMD4A* locus were associated with plasma Ab42/Ab40 ratio^{35,37}.

A list of major studies conducted over the past decade is reported in **Table 1**, along with the year of completion, the ethnic group investigated, the sample size and the main genetic loci discovered. Most of the variants identified by GWAS are, due to the chip design, in intronic regions³². The absence of known direct effects on the genes makes it difficult to interpret how these variants contribute to AD etiology. Albeit GWAS variants located outside of coding regions may play regulatory roles (i.e. altering expression, DNA methylation, or splicing³⁸), GWAS SNPs most often tag the true disease-associated markers through linkage disequilibrium (LD). In order to minimize the risk of false positive findings, replication in independent datasets and functional validation is essential. For replication, several issues need to be considered: as described above, it is common to observe different loci associated with the same disease across independent samples, especially if derived from different ethnic groups. This phenomenon (“allelic heterogeneity”) is expected in complex diseases because of the occurrence of pathogenic mutations across multiple domains of disease genes. Also common is the absence of a specific variant in some datasets or ethnic groups (“locus heterogeneity”), as are differences in LD patterns which depend on the ancestral background, and differences in allele frequencies. Several studies have made an effort to examine transcription and splicing of the genes likely tagged by identified GWAS variants. Expression of a *CLU* isoform influenced by the GWAS variant rs11136000 increases *CLU* protein secretion in persons with high burden of AD neuropathology³⁹. In addition, increased expression of CD33 is associated with AD and inhibits

microglial A β uptake^{40,41}, and copy number variants (CNVs) located within the *CRI* gene are associated with AD⁴².

Rare variants associated with AD identified by next generation sequencing.

Rare genetic variants (i.e. those that occur in less than 1% of the world's population) may be disproportionately important in complex diseases such as AD. In fact, studies have suggested that genes containing common variants with modest effects on complex traits are likely to also contain rare variants with larger effects⁴³. The recent advances in massively-parallel next generation sequencing techniques enable the rapid screen of entire genomes in a timely manner and several rare and low-frequency functional variants conferring strong effects on AD risk have been recently discovered. Although these variants are expected to display large effect sizes on disease risk, due to their rarity the power calculation is a fundamental step in designing the study. A particularly powerful method to soundly implicate rare or low frequency variants by sequencing is the demonstration of segregation with disease status throughout a multiplex pedigree. In samples of independent cases and controls, which cannot capitalize on demonstration of co-segregation and are thus fully dependent on statistical power, burden tests which collapse rare variants in a defined genomic region into a single variable, combined with bioinformatics annotation, can be highly valuable. However, due to a high error rate inherent to this technology (approximately one incorrect base call per 100–1,000 sequenced nucleotides which can result from each step of the sequencing process e.g. miscalling of bases causing misalignment of short reads and erroneous genome assembly) replication of variants in independent datasets and functional validation is also here highly critical.

APP. A recent whole genome sequence (WGS) effort in 1,795 individuals from Iceland identified a nonsynonymous protective mutation in the *APP* gene (A673T)⁴⁴. This variant decreases β -secretase processing of APP, leading to lower A β burden; clinically, it seems to protect elderly without cognitive impairment against cognitive decline and AD. Follow up studies in elderly from the United States suggested that this variant is extremely rare and might be primarily present in Icelandic and Scandinavian populations.⁴⁵⁻⁴⁸

Rare variants at GWAS loci. Several independent studies identified rare disease-associated variants at loci identified by GWAS (**Table 1**).

1. **TREM2.** In 2013, two back-to-back papers^{49,50} implicated a rare missense variant (rs75932628) predicted to cause a R47H substitution in *TREM2* encoding the triggering receptor expressed on myeloid cells 2 protein. *TREM2* activates immune responses in macrophages and dendritic cells and is involved in chronic inflammation by triggering the production of constitutive inflammatory cytokines. Expression of *TREM2* correlates with cortical A β levels and compromised function is believed to negatively affect A β clearance. A prompting effect of *TREM2* on cytokine levels might trigger an inflammatory response and neuronal death. Independent sequencing studies in various populations confirmed the initial findings, in part with variants within different regions of the gene conferring AD risk.⁵¹⁻⁵⁶ R47H has also been associated with history of parental late-onset AD, earlier maternal age-at-onset (AAO),⁵⁷ and early-onset AD.⁵⁸ We further confirmed this finding in a sample of 5,896 African Americans⁵⁴: while the rs75932628 did not pass quality control because of low minor allele frequency, five other SNPs and the gene-based test were significantly associated with AD.

2. ***ABCA7; BIN1; CD2AP; EPHA1***. In a recent study by Vardarajan et al. conducted in three independent AD cohorts (176 patients from 124 Caribbean Hispanics families, 120 patients and 33 unaffected individuals from the 129 NIA-LOAD Family Study; and 263 unrelated Canadian individuals of European ancestry),⁵⁹ targeted sequencing of the GWAS loci revealed an excess burden of deleterious coding mutations in *ABCA7* and *BIN1*. Specifically, mutations in *ABCA7* and *CD2AP* were highly significant in Caucasian cases, whereas mutations in *EPHA1* and *BIN1* were significant in Caribbean Hispanics. Furthermore, the *EPHA1* variant segregated completely in an extended Caribbean Hispanic family and was also replicated in the Caucasians. This finding of causative sequence variants in *BIN1* was confirmed in a cohort of Han Chinese.⁶⁰

3. ***PLD3***. In a WES study in 14 multiplex LOAD families, Cruchaga et al.⁶¹ identified a rare genetic variant (Val232Met) in the gene encoding phospholipase D3 (*PLD3*) that segregated with disease status in two families and doubled risk for AD in several independent case control datasets. Burden analyses further suggested a deleterious sum effect of multiple variants across the gene. *PLD3* belongs to the family of PLD proteins which catalyze the hydrolysis of phosphatidylcholine resulting in free choline and phosphatidic acid. Phosphatidic acid and its derivatives are involved in several pathways such as cytoskeleton formation, membrane trafficking and receptor-mediated endocytosis and exocytosis in addition to neuronal growth and differentiation. In addition, there is evidence for interactions of PLD family members with components of the amyloid cascade⁶². In the study by Cruchaga et al.⁶¹ *PLD3* overexpression significantly reduced APP, A β 42 and A β 40, while knockdown of *PLD3* had inverse effects. Several

sequencing studies, that sought to replicate the findings by Cruchaga et al. failed⁶³⁻⁶⁵ or observed only marginal significance⁶⁶. While several factors, in particular low statistical power and population stratification may explain these discrepancies, additional data are required to clarify the role of rare variants in *PLD3* on AD risk.

4. ***SORL1***. *SORL1* (sortilin-related receptor, L[DLR class] A-type repeats containing) is one of five members of the vacuolar protein sorting-10 domain-containing receptor family. *SORL1* has been clearly implicated in retromer-mediated intracellular trafficking of APP out of the endosomes into the recycling pathway⁶⁷ thereby redirecting APP away from the endosome-lysosome pathway in which it is cleaved into neurotoxic A β 40 and 42 fragments. Following the identification of *SORL1* as an AD susceptibility gene in candidate gene studies and GWAS in various ethnic groups, a recent family- and cohort-based sequencing study in Caribbean Hispanic multiplex families and cases and controls, and a cohort of non-Hispanic White cases and controls identified two AD-associated rare (rs117260922-E270K and rs143571823-T947M) and one common variant (rs2298813-A528T) in *SORL1*⁶⁸. Transfected cell lines for all three variants showed increased A β 40 and A β 42 secretion, and all variants increased the level of APP at the cell surface, in line with a deleterious impact on APP sorting into the recycling pathway⁶⁸.
5. ***MAPT* and *GRN***. Mutations in *MAPT* (encoding the microtubule-associated protein tau) and *GRN* (encoding progranulin) are major causes of frontotemporal lobar degeneration (FTLD). Sequence variants in both genes have been associated with sporadic and familial LOAD.⁶⁹⁻⁷³ While in some of these cases there is clinical evidence of underlying AD

pathology, such as a positive PIB-PET scan, a CSF profile consistent with AD, or temporoparietal atrophy on MRI, in almost all of these studies autopsy confirmation of the clinical diagnosis was lacking. As a consequence, it has been repeatedly suggested that these mutation carriers with a clinical AD diagnosis might in fact have mixed FTLN and AD neuropathology, and that genetic variants in these genes are restricted to the FTLN spectrum.^{70,73} Genomic studies with clinicopathologic data are needed to clarify this issue.

6. Rare variants in other genes. Additional studies have reported rare variants in the cholesterol and phospholipid regulator, ATP-binding cassette transporter, *ABCA1*, nicastrin⁷⁴, the A-kinase anchor protein 9 (*AKAP9*)⁷⁵, *NOTCH3*⁷⁶ and the alpha-secretase Disintegrin and metalloproteinase domain-containing protein 10 (*ADAM10*)^{77,78}. Also these associations need to be validated by further studies.

Structural variants associated with AD. There is increasing evidence that also structural variants such as deletions, duplications, insertions, inversions or translocations contribute substantially to the etiology of complex diseases. Due to the complexity of sequence alterations caused, the mechanistic links between structural variants and phenotypes of interest are often difficult to assess and replicate.

Copy number variants (CNV) analysis in the *APP* locus revealed a fully penetrant 380-kb duplication in a family⁷⁹ and another CNVs was described in *CRI*⁴². In 2014, Hooli et al. reported 10 novel CNVs across different chromosomes in 261 early-onset familial AD cases⁸⁰. A ~470 kb duplication was identified on chromosome 15q11.2 encompassing five genes

(*TUBGCP5*, *CYFIP1*, *NIPA2*, *NIPA1*, and *WHAMML1*)⁸¹. One frameshift mutation (Glu709fs) in *ABCA7* showed a founder effect in a Belgium cohort of 772 cases and 757 controls, and was found to segregate with disease in a family with autosomal dominant inheritance of AD. However, the analysis of structural variants is a rapidly developing field and continues to be an important area of study.

DISCUSSION

The recent advances in systems biology have significantly transformed the field of modern health care from symptom-based disease diagnosis and treatment to the possibility of precision medicine in which patients are optimally treated based on their individual molecular characteristics. The development of high-throughput technologies such as next generation sequencing in combination with sophisticated bioinformatics software for data analysis and visualization has enabled scientists and clinicians to examine whole genomes, transcriptomes, proteinomes and metabolomes in unprecedented detail.

GWAS have now mapped more than 20 common variants associated with AD. In addition, recent NGS studies have led to the identification of several rare functional variants exerting large effects on AD risk. The conducted studies indicate significant locus heterogeneity; in addition, they clearly indicate that not only common variants with small effect sizes, but also rare or low-frequency coding variants with moderate to large effect sizes contribute to AD risk. In summary, the findings of these studies clearly point to specific molecular pathways, in particular immune response and inflammation, cell migration, APP and tau pathology, hippocampal synaptic function, cytoskeletal function and axonal transport, regulation of gene expression and post-translational modification of proteins, and microglial and myeloid cell function.

Despite these recent advances in AD genomics, a significant part of the genetic contribution to AD remains unexplained. The identified common variants outside the *APOE* locus have population attributable risks (PAR) of 1-8%;³² the PAR of the detected rare variants are - by definition – even smaller. Additional common variants could be identified by GWAS if the sample sizes are increased, additional ethnic groups are studied or 'borderline' associations (such as those with $5 \times 10^{-8} > p\text{-value} \leq 1 \times 10^{-7}$) are more closely investigated. The identification of rare variants remains challenging due to several factors including population stratification and low statistical power due to low minor allele frequencies. Potential solutions here are studies in highly selected populations at increased risk for disease such as multiplex families, extreme-trait designs such as focusing on early-onset AD or subjects at the extreme ends of memory function, and investigating different ethnic populations.

Combined with the significant advances in other high-throughput sequencing strategies, such as ChIP-seq, or Chem-seq, analysis of cell free RNA, proteomics or metabolomics, advanced analysis methods such as haplotype resolving strategies or analysis of non-coding RNA, and functional studies, 'omics' information in unprecedented detail can now be generated, providing scientists with the ability to gain a comprehensive view of molecular disease pathways. This characterization of molecular networks is expected to significantly increase treatment efficacy and reduce adverse effects of available drugs and identify novel targets for treatment and prevention. With regard to personal genomics, it is anticipated that specific enough information about AD risk can be gained to influence lifestyle choices and employ non-invasive monitoring

programs such as brain imaging techniques, develop effective genetic testing methods, and design effective personalized therapies.

Acknowledgements:

Dr. Tosto was supported by NIH grants 2U01AG032984-06, 2P50AG008702-26, 2RF1AG015473-16A1 and Department of Defense grant W81XWH.

Dr. Reitz was supported by NIH grants UF1AG047133, P50 AG08702, U01AG032984 and RF1AG015473 and Department of Defense grant W81XWH.

REFERENCES

1. Mayeux R. Epidemiology of neurodegeneration. *Annual review of neuroscience*. 2003;26:81-104.
2. Tosto G, Monsell SE, Hawes SE, Bruno G, Mayeux R. Progression of Extrapyrarnidal Signs in Alzheimer's Disease: Clinical and Neuropathological Correlates. *J Alzheimers Dis*. 2015;49(4):1085-1093.
3. Goate A, Chartier-Harlin MC, Mullan M, et al. Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease. *Nature*. Feb 21 1991;349(6311):704-706.
4. Levy-Lahad E, Wasco W, Poorkaj P, et al. Candidate gene for the chromosome 1 familial Alzheimer's disease locus. *Science*. Aug 18 1995;269(5226):973-977.
5. Sherrington R, Rogaev EI, Liang Y, et al. Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's disease. *Nature*. Jun 29 1995;375(6534):754-760.
6. Corder EH, Saunders AM, Strittmatter WJ, et al. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science*. Aug 13 1993;261(5123):921-923.
7. Farrer LA, Cupples LA, Haines JL, et al. Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease. A meta-analysis. APOE and Alzheimer Disease Meta Analysis Consortium. *JAMA*. Oct 22-29 1997;278(16):1349-1356.
8. Slooter AJ, Cruts M, Kalmijn S, et al. Risk estimates of dementia by apolipoprotein E genotypes from a population-based incidence study: the Rotterdam Study. *Arch Neurol*. Jul 1998;55(7):964-968.
9. Ashford JW, Mortimer JA. Non-familial Alzheimer's disease is mainly due to genetic factors. *J Alzheimers Dis*. Jun 2002;4(3):169-177.
10. Breitner JC, Wyse BW, Anthony JC, et al. APOE-epsilon4 count predicts age when prevalence of AD increases, then declines: the Cache County Study. *Neurology*. Jul 22 1999;53(2):321-331.
11. Gomez-Isla T, West HL, Rebeck GW, et al. Clinical and pathological correlates of apolipoprotein E epsilon 4 in Alzheimer's disease. *Ann Neurol*. Jan 1996;39(1):62-70.
12. Holmes C, Levy R, McLoughlin DM, Powell JF, Lovestone S. Apolipoprotein E: non-cognitive symptoms and cognitive decline in late onset Alzheimer's disease. *J Neurol Neurosurg Psychiatry*. Dec 1996;61(6):580-583.
13. Kurz A, Altland K, Lautenschlager N, et al. Apolipoprotein E type 4 allele and Alzheimer's disease: effect on age at onset and relative risk in different age groups. *J Neurol*. Jun 1996;243(6):452-456.
14. Bennet AM, Di Angelantonio E, Ye Z, et al. Association of apolipoprotein E genotypes with lipid levels and coronary risk. *JAMA*. Sep 19 2007;298(11):1300-1311.
15. Chen J, Li Q, Wang J. Topology of human apolipoprotein E3 uniquely regulates its diverse biological functions. *Proc Natl Acad Sci U S A*. Sep 6 2011;108(36):14813-14818.
16. Schmechel DE, Saunders AM, Strittmatter WJ, et al. Increased amyloid beta-peptide deposition in cerebral cortex as a consequence of apolipoprotein E genotype in late-onset Alzheimer disease. *Proc Natl Acad Sci U S A*. Oct 15 1993;90(20):9649-9653.
17. Rebeck GW, Reiter JS, Strickland DK, Hyman BT. Apolipoprotein E in sporadic Alzheimer's disease: allelic variation and receptor interactions. *Neuron*. Oct 1993;11(4):575-580.
18. Strittmatter WJ, Weisgraber KH, Huang DY, et al. Binding of human apolipoprotein E to synthetic amyloid beta peptide: isoform-specific effects and implications for late-onset Alzheimer disease. *Proc Natl Acad Sci U S A*. Sep 1 1993;90(17):8098-8102.

19. Ma J, Yee A, Brewer HB, Jr., Das S, Potter H. Amyloid-associated proteins alpha 1-antichymotrypsin and apolipoprotein E promote assembly of Alzheimer beta-protein into filaments. *Nature*. Nov 3 1994;372(6501):92-94.
20. Sanan DA, Weisgraber KH, Russell SJ, et al. Apolipoprotein E associates with beta amyloid peptide of Alzheimer's disease to form novel monofibrils. Isoform apoE4 associates more efficiently than apoE3. *J Clin Invest*. Aug 1994;94(2):860-869.
21. Bales KR, Verina T, Cummins DJ, et al. Apolipoprotein E is essential for amyloid deposition in the APP(V717F) transgenic mouse model of Alzheimer's disease. *Proc Natl Acad Sci U S A*. Dec 21 1999;96(26):15233-15238.
22. Bales KR, Verina T, Dodel RC, et al. Lack of apolipoprotein E dramatically reduces amyloid beta-peptide deposition. *Nat Genet*. Nov 1997;17(3):263-264.
23. Verghese PB, Castellano JM, Garai K, et al. ApoE influences amyloid-beta (Abeta) clearance despite minimal apoE/Abeta association in physiological conditions. *Proc Natl Acad Sci U S A*. May 7 2013;110(19):E1807-1816.
24. Harris FM, Brecht WJ, Xu Q, et al. Carboxyl-terminal-truncated apolipoprotein E4 causes Alzheimer's disease-like neurodegeneration and behavioral deficits in transgenic mice. *Proc Natl Acad Sci U S A*. Sep 16 2003;100(19):10966-10971.
25. Tang MX, Stern Y, Marder K, et al. The APOE-epsilon4 allele and the risk of Alzheimer disease among African Americans, whites, and Hispanics. *JAMA*. Mar 11 1998;279(10):751-755.
26. Romas SN, Santana V, Williamson J, et al. Familial Alzheimer disease among Caribbean Hispanics: a reexamination of its association with APOE. *Arch Neurol*. Jan 2002;59(1):87-91.
27. Naj AC, Jun G, Beecham GW, et al. Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. *Nat Genet*. May 2011;43(5):436-441.
28. Harold D, Abraham R, Hollingworth P, et al. Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nat Genet*. Oct 2009;41(10):1088-1093.
29. Lambert JC, Heath S, Even G, et al. Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. *Nat Genet*. Oct 2009;41(10):1094-1099.
30. Seshadri S, Fitzpatrick AL, Ikram MA, et al. Genome-wide analysis of genetic loci associated with Alzheimer disease. *JAMA*. May 12 2010;303(18):1832-1840.
31. Hollingworth P, Harold D, Sims R, et al. Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. *Nat Genet*. May 2011;43(5):429-435.
32. Lambert JC, Ibrahim-Verbaas CA, Harold D, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet*. Dec 2013;45(12):1452-1458.
33. Rogaeva E, Meng Y, Lee JH, et al. The neuronal sortilin-related receptor SORL1 is genetically associated with Alzheimer disease. *Nat Genet*. Feb 2007;39(2):168-177.
34. Reitz C, Jun G, Naj A, et al. Variants in the ATP-binding cassette transporter (ABCA7), apolipoprotein E 4, and the risk of late-onset Alzheimer disease in African Americans. *JAMA*. Apr 10 2013;309(14):1483-1492.
35. Tosto G, Fu H, Vardarajan BN, et al. F-box/LRR-repeat protein 7 is genetically associated with Alzheimer's disease. *Ann Clin Transl Neurol*. Aug 2015;2(8):810-820.
36. Wang HZ, Bi R, Hu QX, et al. Validating GWAS-Identified Risk Loci for Alzheimer's Disease in Han Chinese Populations. *Mol Neurobiol*. Jan 2016;53(1):379-390.
37. Lambert JC, Grenier-Boley B, Harold D, et al. Genome-wide haplotype association study identifies the FRMD4A gene as a risk locus for Alzheimer's disease. *Mol Psychiatry*. Apr 2013;18(4):461-470.

38. Ward LD, Kellis M. Interpreting noncoding genetic variation in complex traits and human disease. *Nat Biotechnol*. Nov 2012;30(11):1095-1106.
39. Ling IF, Bhongsatiern J, Simpson JF, Fardo DW, Estus S. Genetics of clusterin isoform expression and Alzheimer's disease risk. *PLoS One*. 2012;7(4):e33923.
40. Karch CM, Jeng AT, Nowotny P, Cady J, Cruchaga C, Goate AM. Expression of novel Alzheimer's disease risk genes in control and Alzheimer's disease brains. *PLoS One*. 2012;7(11):e50976.
41. Griciuc A, Serrano-Pozo A, Parrado AR, et al. Alzheimer's disease risk gene CD33 inhibits microglial uptake of amyloid beta. *Neuron*. May 22 2013;78(4):631-643.
42. Brouwers N, Van Cauwenberghe C, Engelborghs S, et al. Alzheimer risk associated with a copy number variation in the complement receptor 1 increasing C3b/C4b binding sites. *Mol Psychiatry*. Feb 2012;17(2):223-233.
43. Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. *Nature*. Oct 8 2009;461(7265):747-753.
44. Jonsson T, Atwal JK, Steinberg S, et al. A mutation in APP protects against Alzheimer's disease and age-related cognitive decline. *Nature*. Aug 2 2012;488(7409):96-99.
45. Bamne MN, Demirci FY, Berman S, et al. Investigation of an amyloid precursor protein protective mutation (A673T) in a North American case-control sample of late-onset Alzheimer's disease. *Neurobiol Aging*. Jul 2014;35(7):1779 e1715-1776.
46. Wang LS, Naj AC, Graham RR, et al. Rarity of the Alzheimer disease-protective APP A673T variant in the United States. *JAMA Neurol*. Feb 2015;72(2):209-216.
47. Liu YW, He YH, Zhang YX, et al. Absence of A673T variant in APP gene indicates an alternative protective mechanism contributing to longevity in Chinese individuals. *Neurobiol Aging*. Apr 2014;35(4):935 e911-932.
48. Ting SK, Chong MS, Kandiah N, et al. Absence of A673T amyloid-beta precursor protein variant in Alzheimer's disease and other neurological diseases. *Neurobiol Aging*. Oct 2013;34(10):2441 e2447-2448.
49. Jonsson T, Stefansson H, Steinberg S, et al. Variant of TREM2 associated with the risk of Alzheimer's disease. *N Engl J Med*. Jan 10 2013;368(2):107-116.
50. Guerreiro R, Wojtas A, Bras J, et al. TREM2 variants in Alzheimer's disease. *N Engl J Med*. Jan 10 2013;368(2):117-127.
51. Benitez BA, Jin SC, Guerreiro R, et al. Missense variant in TREML2 protects against Alzheimer's disease. *Neurobiol Aging*. Jun 2014;35(6):1510 e1519-1526.
52. Jin SC, Benitez BA, Karch CM, et al. Coding variants in TREM2 increase risk for Alzheimer's disease. *Hum Mol Genet*. Nov 1 2014;23(21):5838-5846.
53. Jin SC, Carrasquillo MM, Benitez BA, et al. TREM2 is associated with increased risk for Alzheimer's disease in African Americans. *Mol Neurodegener*. 2015;10:19.
54. Reitz C, Mayeux R. TREM2 and neurodegenerative disease. *N Engl J Med*. Oct 17 2013;369(16):1564-1565.
55. Korvatska O, Leverenz JB, Jayadev S, et al. R47H Variant of TREM2 Associated With Alzheimer Disease in a Large Late-Onset Family: Clinical, Genetic, and Neuropathological Study. *JAMA Neurol*. Jun 15 2015.
56. Rosenthal SL, Bamne MN, Wang X, et al. More evidence for association of a rare TREM2 mutation (R47H) with Alzheimer's disease risk. *Neurobiol Aging*. Aug 2015;36(8):2443 e2421-2446.
57. Engelman CD, Kosciuk RL, Jonaitis EM, Hermann BP, La Rue A, Sager MA. Investigation of triggering receptor expressed on myeloid cells 2 variant in the Wisconsin Registry for Alzheimer's Prevention. *Neurobiol Aging*. Jun 2014;35(6):1252-1254.

58. Pottier C, Wallon D, Rousseau S, et al. TREM2 R47H variant as a risk factor for early-onset Alzheimer's disease. *J Alzheimers Dis.* 2013;35(1):45-49.
59. Vardarajan BN, Ghani M, Kahn A, et al. Rare coding mutations identified by sequencing of Alzheimer's disease GWAS loci. *Ann Neurol.* Jun 23 2015.
60. Tan MS, Yu JT, Jiang T, Zhu XC, Guan HS, Tan L. Genetic variation in BIN1 gene and Alzheimer's disease risk in Han Chinese individuals. *Neurobiol Aging.* Jul 2014;35(7):1781 e1781-1788.
61. Cruchaga C, Karch CM, Jin SC, et al. Rare coding variants in the phospholipase D3 gene confer risk for Alzheimer's disease. *Nature.* Jan 23 2014;505(7484):550-+.
62. Oliveira TG, Di Paolo G. Phospholipase D in brain function and Alzheimer's disease. *Biochim Biophys Acta.* Aug 2010;1801(8):799-805.
63. Hooli BV, Lill CM, Mullin K, et al. PLD3 gene variants and Alzheimer's disease. *Nature.* Apr 2 2015;520(7545):E7-8.
64. Heilmann S, Drichel D, Clarimon J, et al. PLD3 in non-familial Alzheimer's disease. *Nature.* Apr 2 2015;520(7545):E3-5.
65. Lambert JC, Grenier-Boley B, Bellenguez C, et al. PLD3 and sporadic Alzheimer's disease risk. *Nature.* Apr 2 2015;520(7545):E1.
66. van der Lee SJ, Holstege H, Wong TH, et al. PLD3 variants in population studies. *Nature.* Apr 2 2015;520(7545):E2-3.
67. Lintzel J, Franke I, Riedel IB, Schaller HC, Hampe W. Characterization of the VPS10 domain of SorLA/LR11 as binding site for the neuropeptide HA. *Biol Chem.* Nov 2002;383(11):1727-1733.
68. Vardarajan BN, Zhang Y, Lee JH, et al. Coding mutations in SORL1 and Alzheimer disease. *Ann Neurol.* Feb 2015;77(2):215-227.
69. Coppola G, Chinnathambi S, Lee JJ, et al. Evidence for a role of the rare p.A152T variant in MAPT in increasing the risk for FTD-spectrum and Alzheimer's diseases. *Hum Mol Genet.* Aug 1 2012;21(15):3500-3512.
70. Cruchaga C, Haller G, Chakraverty S, et al. Rare variants in APP, PSEN1 and PSEN2 increase risk for AD in late-onset Alzheimer's disease families. *PLoS One.* 2012;7(2):e31039.
71. Gerrish A, Russo G, Richards A, et al. The role of variation at AbetaPP, PSEN1, PSEN2, and MAPT in late onset Alzheimer's disease. *J Alzheimers Dis.* 2012;28(2):377-387.
72. Jin SC, Pastor P, Cooper B, et al. Pooled-DNA sequencing identifies novel causative variants in PSEN1, GRN and MAPT in a clinical early-onset and familial Alzheimer's disease Ibero-American cohort. *Alzheimers Res Ther.* 2012;4(4):34.
73. Lee SE, Tartaglia MC, Yener G, et al. Neurodegenerative disease phenotypes in carriers of MAPT p.A152T, a risk factor for frontotemporal dementia spectrum disorders and Alzheimer disease. *Alzheimer Dis Assoc Disord.* Oct-Dec 2013;27(4):302-309.
74. Lupton MK, Proitsi P, Danillidou M, et al. Deep sequencing of the Nicastrin gene in pooled DNA, the identification of genetic variants that affect risk of Alzheimer's disease. *PLoS One.* 2011;6(2):e17298.
75. Logue MW, Schu M, Vardarajan BN, et al. Two rare AKAP9 variants are associated with Alzheimer's disease in African Americans. *Alzheimers Dement.* Nov 2014;10(6):609-618 e611.
76. Guerreiro RJ, Lohmann E, Kinsella E, et al. Exome sequencing reveals an unexpected genetic cause of disease: NOTCH3 mutation in a Turkish family with Alzheimer's disease. *Neurobiol Aging.* May 2012;33(5):1008 e1017-1023.
77. Cai G, Atzmon G, Naj AC, et al. Evidence against a role for rare ADAM10 mutations in sporadic Alzheimer disease. *Neurobiol Aging.* Feb 2012;33(2):416-417 e413.
78. Kim M, Suh J, Romano D, et al. Potential late-onset Alzheimer's disease-associated mutations in the ADAM10 gene attenuate {alpha}-secretase activity. *Hum Mol Genet.* Oct 15 2009;18(20):3987-3996.

79. Hooli BV, Mohapatra G, Mattheisen M, et al. Role of common and rare APP DNA sequence variants in Alzheimer disease. *Neurology*. Apr 17 2012;78(16):1250-1257.
80. Hooli BV, Kovacs-Vajna ZM, Mullin K, et al. Rare autosomal copy number variations in early-onset familial Alzheimer's disease. *Mol Psychiatry*. Jun 2014;19(6):676-681.
81. Ghani M, Pinto D, Lee JH, et al. Genome-wide survey of large rare copy number variants in Alzheimer's disease among Caribbean hispanics. *G3 (Bethesda)*. Jan 2012;2(1):71-78.

Table 1. Major Alzheimer's disease genome-wide association studies (GWAS) and whole-exome/whole-genome sequencing studies (WES/WGS) performed

	Year	Ethnic group	Sample size*	Genes identified
GWAS				
Lambert et al.²⁹	2009	Caucasian	2,032 cases; 5,328 controls	<i>CLU, CR1</i>
Harold et al.²⁸	2009	Caucasian	3,941 cases; 7,848 controls	<i>CLU, PICALM</i>
Seshadri et al.³⁰	2010	Caucasian	3,006 cases; 14,642 controls	<i>BIN1, XOC3L2/BLOC1S3/MARK4, CLU, PICALM</i>
Naj et al.²⁷	2011	Caucasian	8,309 cases; 7,366 controls	<i>MS4A4A, CD2AP, CD33, EPHA1, CR1, CLU, BIN1, PICALM</i>
Hollingworth et al.³¹	2011	Caucasian	6,688 cases; 13,685 controls	<i>ABCA7, MS4A6A/MS4A4E, EPHA1, CD33, CD2AP</i>
Lambert et al.³²	2013	Caucasian	17,008 cases; 37,154 controls	<i>CR1, BIN1, CD2AP, EPHA1, CLU, MS4A4A, PICALM, ABCA7, HLA-region, PTK2B, SORL1, SLC24A4-RIN3, INPP5D, MEF2C, NME8, ZCWPW1, CELF1, FERMT2, CASS4</i>
Lambert et al.³⁷	2013	Caucasian	2025 cases; 5328 controls	<i>FRMD4A</i>
Reitz et al.³⁴	2013	African American	1,968 AD cases; 3,928 controls	<i>ABCA7</i> , intergenic locus on 5q35.2
Tosto et al.³⁵	2015	Caribbean Hispanics	3,001 cases; 2,299 controls	<i>FBXL7</i>
WES/WGS				
Jonsson et al. Guerreiro et al.^{49,50}	2013	Caucasian	3550 cases; 8,888 controls 1,092 cases; 1,107 controls	<i>TREM2</i>
Vardarajan et al.⁵⁹	2015	Caucasian, Caribbean Hispanics	330 cases; 86 controls (Caucasians) + 176 Caribbean Hispanic cases	<i>ABCA7, BIN1, CD2AP, EPHA1</i>
Vardarajan et al.⁶⁸	2014	Caucasian, Caribbean Hispanics	154 cases; 80 controls (Hispanics) + 211 Caucasian cases	<i>SORL1</i>
Cruchaga et al.⁶¹	2014	Caucasian	14 large affected families	<i>PLD3</i>

*number refers to discovery datasets.