



# Genomics and Functional Genomics of Alzheimer's Disease

M. Ilyas Kamboh<sup>1</sup>

Accepted: 25 October 2021 / Published online: 21 December 2021  
© The Author(s) 2021

## Abstract

Alzheimer's disease (AD) is a complex and multifactorial neurodegenerative disease. Due to its long clinical course and lack of an effective treatment, AD has become a major public health problem in the USA and worldwide. Due to variation in age-at-onset, AD is classified into early-onset (< 60 years) and late-onset ( $\geq$  60 years) forms with early-onset accounting for only 5–10% of all cases. With the exception of a small number of early-onset cases that are afflicted because of high penetrant single gene mutations in *APP*, *PSEN1*, and *PSEN2* genes, AD is genetically heterogeneous, especially the late-onset form having a polygenic or oligogenic risk inheritance. Since the identification of *APOE* as the most significant risk factor for late-onset AD in 1993, the path to the discovery of additional AD risk genes had been arduous until 2009 when the use of large genome-wide association studies opened up the discovery gateways that led the identification of ~95 additional risk loci from 2009 to early 2022. This article reviews the history of AD genetics followed by the potential molecular pathways and recent application of functional genomics methods to identify the causal AD gene(s) among the many genes that reside within a single locus. The ultimate goal of integrating genomics and functional genomics is to discover novel pathways underlying the AD pathobiology in order to identify drug targets for the therapeutic treatment of this heterogeneous disorder.

**Keywords** Alzheimer's disease · SNPs · Genomics · Genome-wide association studies · Functional genomics

## Introduction

Alzheimer's disease (AD) is a complex and multifactorial neurodegenerative disease and the leading cause of dementia among elderly, accounting for 60–80% of all dementia cases. However, a small number of people (5–10%) develop AD at a younger age, and due to this variation in age-at-onset (AAO), the disease is classified into early-onset (EOAD) with AAO < 60 years and late-onset (LOAD) with AAO  $\geq$  60 years. Deposition of amyloid-beta ( $A\beta$ ) plaques and the formation of neurofibrillary tangles (tau pathology) in the brain are required for the neuropathological diagnosis of AD [1]. Approximately half of the cases with AD dementia have solely AD brain pathology, and the remaining cases have coexisting pathological brain changes of AD and other dementia(s), like vascular dementia and Lewy body dementia. Such subjects with mixed pathologies are termed having mixed dementia [2]. Due to its long clinical course,

transitioning from mild to moderate and severe stage, AD is a major public health problem in the USA and worldwide. On average, a person with AD can survive 4–8 years after the diagnosis, but some can survive as long as 20 years. There are over 50 million people worldwide living with AD or dementia, and this number is projected to increase 152 million in 2050 [3]. Currently, there are 6.2 million AD cases in the USA [2]. The estimated annual cost for caring of AD patients in the USA is projected to be \$1.1 trillion in 2050 when the number of AD cases would reach to ~14 million, if no medical breakthroughs are found.

Among the multiple known risk factors for AD, the strongest evidence is for age, biological (gender), and genetic differences. Although AD is not a normal part of aging, older age is the greatest risk factor for AD. The prevalence of AD increases dramatically after age 60: 5% in age group 65–74, 14% in age group 75–85, and 35% in age group 85 and older [2]. Currently, about 58 million Americans are age 65 and older and the Bureau of the Census estimates that this number will be 58 million by the year 2050. Due to this alarming increase in the elderly population and with the possibility that a large fraction of this elderly population would suffer from AD, it is essential to understand the causes of AD so that effective preventative measures could

✉ M. Ilyas Kamboh  
kamboh@pitt.edu

<sup>1</sup> Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA

be devised. In fact, the oldest members of the baby-boom generation of Americans (born between 1946 and 1964) have already turned age 75 in 2021. The prevalence of AD is higher in women than men; almost two-thirds of the AD cases are women [2]. The estimated lifetime risk for AD among Americans at age 45 or 65 is 20% in women and 10% in men [4]. In addition to the biological difference in men and women, this gender difference could be due to “survival effect” (women live longer than men and the older age is the greatest risk for AD) or “survival bias” (men who survive beyond age 60 have healthier cardiovascular profile than women have and thus have lower risk of AD) [2].

The role of genetics in the etiology is well recognized in both EOAD and LOAD. Mutations in three casual genes for EOAD have been identified (*APP*, *PSEN1*, and *PSEN2*) that follow the autosomal dominant inheritance pattern. LOAD that comprises 90–95% of the cases is genetically more complex with heritability estimates from 58 to 70% [5, 6]. *APOE* was identified as the first susceptibility gene for LOAD in 1993, which was followed by a plethora of linkage, and positional and biological candidate gene approaches to identify additional genes but without a success [7, 8]. Until 2009, *APOE* was the only established susceptibility gene for LOAD. Since 2009, substantial progress has been made in LOAD genetics via large-scale case–control genome-wide association studies (GWAS) as well as meta-analyses of GWAS that have led to the identification numerous susceptibility loci. Recent applications of whole-exome microarray, whole-exome sequencing (WES), and whole-genome sequencing (WGS) have also identified rare variants in additional novel LOAD genes. However, the common single nucleotide polymorphism (SNP)-based heritability from known AD genes ranges from 13 to 33% [9–11], indicating that much of the AD genetic variance remains unexplained. One of the likely explanations of relatively low genetic variance observed with known LOAD loci in case–control studies is the inclusion of “non-AD dementia” subjects in the case group based on clinical AD diagnosis that is prone to high misclassification, and the control group may include many “presumptive AD cases” because the preclinical stage of AD begins decades before the initial clinical symptoms of AD [12, 13]. In order to overcome this misclassification and to better understand the underlying mechanisms of AD clinical expression, the 2018 research framework proposed by the National Institute on Aging and Alzheimer's Association defines AD as a biological construct that is identified by biomarkers in living people [14].

The purpose of this review is to provide an update on the genomics landscape of AD and then discuss the ongoing efforts in integrating GWAS findings in LOAD with functional genomics in order to identify putative functional genes and molecular pathways.

## Early-Onset Alzheimer's Disease

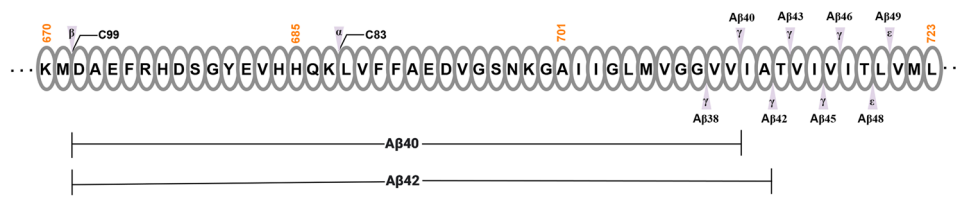
Approximately 5–10% of all AD cases are classified as EOAD, of which 35–65% are familial (FEOAD) and the remaining are sporadic [15, 16]. Of the FEOAD cases, about 10–15% follow an autosomal dominant inheritance pattern [17] due to mutations in three known genes: *APP*, *PSEN1*, and *PSEN2*. Approximately 90% of the remaining FEOAD cases are most likely due to autosomal recessive inheritance pattern [6]. Mutations in the three-causal EOAD genes regulate the production, aggregation, and degradation of beta-amyloid peptides (A $\beta$ s), which are primary components of amyloid plaques.

A $\beta$ s are derived from the amyloid precursor protein (APP), which is a transmembrane protein consisting of an amino-terminal domain of 699 amino acids (in the longest APP-770) in the extracellular or intraluminal region, a short transmembrane region of 24 amino acids (from 700 to 723), and a 47-amino acid long intracellular domain known as AICD (APP intracellular domain). APP undergoes three alternative proteolytic cleavages by  $\alpha$ -,  $\beta$ -, and  $\gamma$ -secretases within or surrounding the A $\beta$  sequence to generate either smaller fragments in the non-amyloidogenic pathway or longer fragments in the amyloidogenic pathways [18]. APP is first cleaved by either  $\alpha$ -secretase at amino acid 687 or  $\beta$ -secretase (BACE1) at amino acid 671 to generate a transmembrane carboxyl fragment of APP-C83 or APP-C99, respectively (Fig. 1). APP-C83 or APP-C99 is then cleaved by the transmembrane aspartyl protease  $\gamma$ -secretase to generate the smaller harmless peptides or the longer pathogenic peptides, respectively [19]. Initially, APP-C99 is cleaved by  $\gamma$ -secretase through its endopeptidase activity ( $\epsilon$ -site) to generate A $\beta$ 48 or A $\beta$ 49 peptide, which is then further cleaved sequentially by the carboxypeptidase activity of  $\gamma$ -secretase to produce A $\beta$ 45, A $\beta$ 42, and A $\beta$ 38 peptides from A $\beta$ 48 or A $\beta$ 46, A $\beta$ 43, and A $\beta$ 40 peptides from A $\beta$ 49 (Fig. 1). Of these peptides, A $\beta$ 40 is the most common and A $\beta$ 42, which constitutes approximately 10% of the total A $\beta$ , is particularly associated with the development of AD [20]. With the extra two hydrophobic amino acids, A $\beta$ 42 is more prone to aggregation than A $\beta$ 40 and is found abundantly in neuritic plaques and amyloid-bearing microvessels.

Human  $\gamma$ -secretase complex consists of four protein subunits: presenilin (PSEN), presenilin enhancer 2 (PEN2), anterior pharynx-defective-1 (APH1), and nicastrin (NCT). The catalytic activity of  $\gamma$ -secretase is determined by PSEN, which exists in two isoforms (PSEN1 and PSEN2). APH1 stabilizes the complex and has two isoforms (APH1A and APH1B), PEN2 is essential for  $\gamma$ -secretase maturation, and NCT plays a role in substrate binding [19–22].

A large number of high-penetrant mutations have been identified in the *APP*, *PSEN1*, and *PSEN2* genes that

### β-Amyloid (Aβ) Peptides



**Fig. 1** Generation of β-amyloid (Aβ) peptides. The APP amino acid sequence from position 670 to 723 is shown, including the transmembrane region from amino acid 701 to 723. The cleavage sites of α-, β-, and γ-secretases are labelled. Following the generation of APP-C99 by β-secretase, it is cleaved by γ-secretase through its endopeptidase activity (ε-site) to generate Aβ48 or Aβ49 peptide, which is then

further cleaved sequentially by the C-terminal peptidase activity of γ-secretase to produce Aβ45, Aβ42, and Aβ38 peptides from Aβ48 (shown at the right bottom end) or Aβ46, Aβ43, and Aβ40 peptides from Aβ49 (shown at the right top end). The full lengths of Aβ40 and Aβ42 peptides are illustrated at the bottom. The start position of APP-C83 following the cleavage with α-secretase is also shown

regulate Aβs production (especially Aβ40 and Aβ42) and cause EOAD. As of October 2021, Alzforum lists 50, 309, and 55 AD-associated mutations in *APP*, *PSEN1*, and *PSEN2*, respectively ([www.alzforum.org/mutations](http://www.alzforum.org/mutations)). Altogether, pathogenic mutations in these three genes account for approximately 7% of the EOAD cases, with a 6% contribution from *PSEN1* and ~1% from the other two genes [23]. The majority of the *APP* mutations are missense that occurs in or around the Aβ sequence and lead to either overproduction of total Aβ or increased Aβ42/Aβ40 ratio. With the exception of two mutations (p.Ala673Val and one amino acid deletion, p.Glu693Δ), which are autosomal recessive, all other *APP* mutations are autosomal dominant [16, 24]. Two de novo *APP* duplications have also been described [23]. Overproduction of Aβ is also a key feature in Down syndrome patients who have three copies of the *APP* containing chromosome 21 and develop a neuropathology that is indistinguishable from AD [24]. These findings tend to support the amyloid cascade hypothesis that deposition of Aβ in the brain is the initiating factor in the pathogenesis of AD [25].

The majority of the *PSEN1* and *PSEN2* mutations are missense with autosomal dominant inheritance, although some autosomal recessive and de novo mutations have also been described [23]. Mutations in the *PSEN1* and *PSEN2* genes increase Aβ42, but decrease Aβ40, resulting in an increased Aβ42/Aβ40 ratio. These mutations also shift Aβ40 and Aβ42 to longer and more harmful species. Increased Aβ42/Aβ40 ratio has also shown to be associated with disease progression, as reflected in early AAO in familial AD. However, a comprehensive in vitro analysis of the effect of 138 *PSEN1* pathogenic mutations on γ-secretase activity in the production of Aβ42 and Aβ40 found no significant correlation between the Aβ42/Aβ40 ratio and the mean AAO [26]. Furthermore, about 90% of the examined 138 *PSEN1* mutations resulted in reduced production of Aβ42 and Aβ40 and 10% of them lead to decreased Aβ42/Aβ40 ratio in the

experimental system used in the study. These findings appear to contradict the amyloid hypothesis and suggest the possible involvement of multiple mechanisms in the etiology of AD where the amyloid pathway may just constitute one of them [26]. This also highlights the need to assess the functional nature of all the *APP*, *PSEN1*, and *PSEN2* mutations because the pathogenicity of most of these mutations is yet to be examined.

To date, no mutations have been described to be associated with EOAD either in the *ADAM10* or *ADAM17* (key components of α-secretase complex) and *BACE1* genes or the genes coding for the other members of the γ-secretase complex (*PEN2*, *APH-1*, *NCT*). However, rare missense mutations in the *SORL1* gene have been suggested to play an important role in subset of FEOAD [27, 28].

### Role of Early-Onset AD Genes in Late-Onset AD

Despite the high penetrance of *APP*, *PSEN1*, and *PSEN2* mutations in EOAD, their role in LOAD susceptibility was not clear, and so was the case with other genes involved in the regulation of APP metabolism and/or Aβ generation. This was changed, however, in 2009 when two rare missense mutations in *ADAM10* (p.Gln170His; p.Arg181Gly) were reported to cosegregate in LOAD families [29]. These mutations attenuated α-secretase activity of ADAM10 and increased Aβ plaques and reactive gliosis in transgenic mice [30]. In 2012, two studies reported rare risk and protective variants of *APP* in LOAD. In the first study, a rare pathogenic *APP* coding variant (p.Asn660Tyr) exhibiting high disease penetrance was reported in multiple members of an LOAD family, which was absent in 1346 controls and 12,481 subjects not enriched for AD [31]. In the second study, a rare but widely distributed *APP* coding variant (p.Ala673Thr/rs63750847) was found to be protective against sporadic LOAD and cognitive impairment in non-AD elderly individuals among Icelanders, Finnish, Norwegian, and Swedish

populations [32]. The minor allele frequency (MAF) of this variant was 0.13% in AD cases compared with 0.45% in population controls and 0.62% in population controls aged 85 or greater. Subsequent large LOAD studies in US Whites found either no or only sporadic examples of p.Ala673Thr [33, 34], indicating its confinement to individuals of the origin from the Nordic countries. This variant is also absent among Chinese [35]. The p.Ala673Thr variant is present at the same site as the autosomal recessive p.Ala673Val pathogenic mutation and is adjacent to the BACE1 cleavage site at position 671. The protective p.Ala673Thr mutation is associated with reduced production of amyloidogenic A $\beta$  peptides by about 40%, and the generated A $\beta$  is less prone to aggregation [32, 35]. This protective variant was also detected in a 105-year-old Finnish demented subject who showed little A $\beta$  pathology [36].

In addition to rare variants, recent large GWAS have identified common variants in the *ADAM10*, *ADAM17*, *APH1B*, and *APP-ADAMTS1* gene regions to be associated with LOAD susceptibility (discussed later), further providing support to the amyloid cascade hypothesis in at least subset of the LOAD patients.

## Late-Onset Alzheimer's Disease

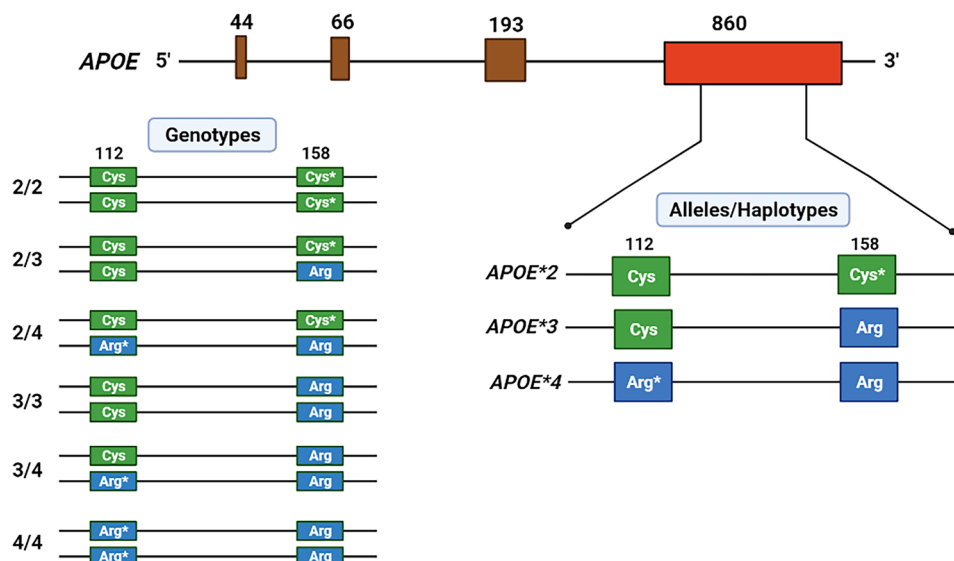
### *APOE* Polymorphism

The strongest risk factor for LOAD is the common three-allele *APOE* polymorphism: *APOE*\*2, *APOE*\*3, and *APOE*\*4 [37, 38], resulting in six genotypes (2/2, 2/3, 2/4, 3/3, 3/4, 4/4). This polymorphism is characterized by missense mutations at the first base of codon 112 and codon 158 (Fig. 2). *APOE*\*3 is the most common allele

and code for Cys (TGC) at position 112 and Arg (CGC) at position 158. *APOE*\*4 and *APOE*\*2 differ from *APOE*\*3 by having Arg (CGC) at position 112 and Cys (TGC) at position 158, respectively. Due to strong linkage disequilibrium (LD) between the two sites, the three alleles also define three rather than the four expected haplotypes: E3 has Cys at position 112 and Arg at position 158, E4 has Arg at both positions, and E2 has Cys at both positions. The expected fourth haplotype having Arg at position 112 (*E*\*4 allele) and Cys at position 158 (*E*\*2 allele) has been observed only rarely [39]. These structural differences in *APOE* alleles have profound effect on the function of the ApoE protein in lipid metabolism/cardiovascular function and in determining AD risk [37, 40–45]. Although *APOE*\*4 is associated with adverse plasma lipid profile (high LDL-cholesterol, high ApoB, low plasma ApoE levels) and high risk of heart disease and AD, the effect of *APOE*\*2 on these traits is in opposite direction. *APOE*\*4 is also associated with AD-related proteinopathies: A $\beta$ , tau,  $\alpha$ -synuclein, and TDP-43 [42, 43]. The differential *APOE* allelic effect on AD risk is likely regulated in large part due to its impact on AD-related proteinopathies. *APOE*\*4 may also have direct pathologic effects on neurons and the blood–brain barrier function independent of its effects on amyloid and tau pathologies [43].

The frequencies of three *APOE* alleles and the proportion of people who carry them (genotype carriers) differ significantly among the major human racial groups (Table 1). The aboriginal populations of Australia and America seem to have no or only sporadic presence of the *APOE*\*2 allele, which is present at an allele frequency of about 4% in Blacks and 8% in most other populations. Likewise, the average occurrence of *APOE*\*4 allele frequency (genotype carriers) is 7% (14%) in Asians–Japanese/Chinese; 14% (25%) in Whites; 26–37% (40–50%) in Blacks, Australian

**Fig. 2** Structure of the *APOE* gene with four exons (top). Non-synonymous mutations at codon 112 (Cys112Arg) and codon 158 (Arg158Cys) in exon 4 code for three common alleles/haplotypes: *APOE*\*2, *APOE*\*3, and *APOE*\*4 (bottom right), resulting in six genotypes (bottom left). The most common *APOE*\*3 allele has Cys at position 112 and Arg at position 158 in the ApoE protein. The amino acid change in *APOE*\*4 at codon 112 is indicated by Arg\*, and the amino acid change in *APOE*\*2 at codon 158 is indicated by Cys\*





**Table 1** *APOE* allele frequencies in major racial groups\*

Population	<i>APOE</i> *2	<i>APOE</i> *3	<i>APOE</i> *4
<b>Whites</b>			
US Whites	0.07	0.79	0.14
Germans	0.08	0.77	0.15
French	0.13	0.74	0.13
Finns	0.04	0.73	0.23
Dutch	0.08	0.75	0.17
<b>Africans</b>			
American blacks	0.03	0.71	0.26
Nigerians	0.03	0.67	0.30
Sudanese	0.08	0.63	0.29
<b>Asians</b>			
Japanese	0.08	0.85	0.07
Chinese	0.10	0.83	0.07
<b>Oceanians</b>			
New Guineans	0.14	0.49	0.37
Australian Aborigines	0.00	0.74	0.26
Polynesians	0.11	0.63	0.26
<b>Native Americans</b>			
Amerindians	0.00	0.82	0.18
Mayans	0.00	0.91	0.09
Eskimos	0.01	0.76	0.23

\*Adapted from Kamboh 1995 [37]

Aborigines, and Pacific populations; and 23% (40%) in Eskimos. Although the prevalence of AD is not well documented in the aboriginal populations of Australia and America as well as in the Pacific populations, these populations seem to be at high risk of developing AD by having a very high frequency of *APOE*\*4 carriers should their life expectancy increases, as seen among the European or European-derived White populations.

*APOE*\*4 is a significant risk factor for AD and shows gene dosage effect on AD risk in populations with diverse racial/ethnic backgrounds. However, *APOE*\*4 is neither necessary nor sufficient to develop AD, indicating the involvement of additional genetic factors that can modify the risk of AD. There is also considerable inter-racial variation in the gene dosage effect of *APOE*\*4 on AD risk, as measured in odds ratio (OR). The ORs with one and two copies *APOE*\*4 in major racial groups are as follows: 3.5 and 14.5 in Whites; 1.1–2.2 and 2.2–5.7 in African Americans; 3.1 and 11.8 in Chinese; and 5.6 and 33.1 in Japanese [38, 46–48]. The relatively higher ORs in one Japanese study need to be confirmed in larger Japanese samples because this study used only 336 AD cases and thus may not be representative of the Japanese population. However, the findings of low ORs in African Americans are intriguing despite that the frequency of *APOE*\*4 is almost twice in Africans compared to Whites. This difference may be due to variation in environmental

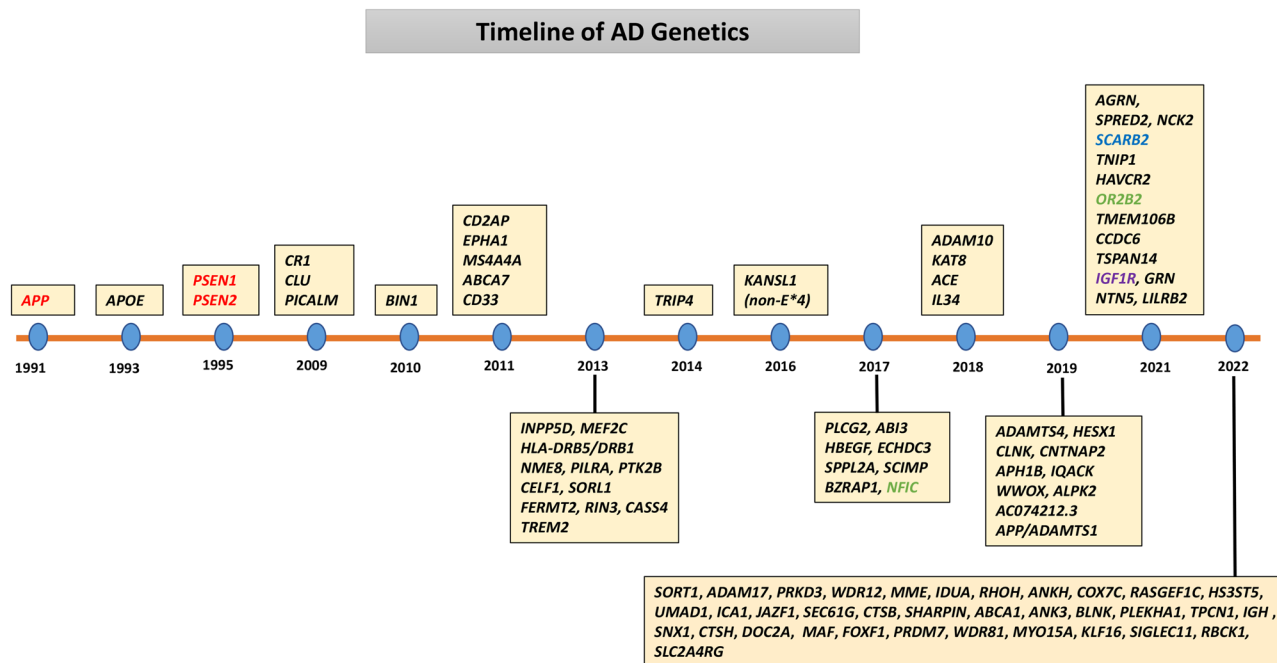
and cultural factors, or it may represent additional genetic variation present in the *APOE* region, which is protective in Africans. Indeed, it has been suggested that the genomic region surrounding *APOE* with African background reduces the risk for *APOE*\*4 carriers [48]. In addition to the common *APOE* polymorphism, which explains about 25% of the genetic variance of AD [9], rare coding and noncoding variants in *APOE* have also been implicated with the risk of AD [49–51].

## Genome-Wide Association Studies

In order to identify additional genes/loci for LOAD, genome-wide association studies (GWAS) were initiated soon after the availability of chip microarrays because this approach is hypothesis free and conceptually would identify all known and unknown genes. However, with the exception of the *APOE* region, no other genome-wide significant (GWS) associations ( $p < 5E - 08$ ) were identified in earlier GWAS published during 2007–2008 [52–59]. The GWS threshold of  $p < 5E - 08$  (0.05/1,000,000) is widely applied to avoid false-positive findings in GWAS comprising approximately 1 million independent association tests. These early negative findings were due to the use of modest sample size and suggested that the effect sizes of yet to be discovered variants were very small as compared with the large effect of *APOE*\*4 and that it would require the use of large number of cases and controls in order to detect small effect sizes. Indeed, the recent use of large GWAS and meta-analyses of GWAS has led to the identification of ~95 loci for LOAD during the period of 2009 to early 2022. The timeline of AD genetics is illustrated in Fig. 3, and the history and outcomes of GWAS are summarized in the below sections and in Table 2. It is worth noting that the gene names given in Fig. 3 and Table 2 represent the closest gene to the lead SNP, not necessarily the causal gene in the locus, and that the gene name used to refer to a locus can change between GWAS depending on the lead variant and its corresponding closest gene. Like other diseases, the focus of the vast majority of genomic studies in LOAD has been on European or European-derived White populations.

## GWAS in Europeans/Whites

In 2009, two large GWAS published back-to-back in *Nature Genetics* identified three novel GWS loci (*CLU*, *PICALM*, and *CR1*) in addition to *APOE* [60, 61]. The combined sample in the first study, which was derived from a collaborative consortium Genetic and Environmental Risk in Alzheimer's Disease (GERAD) from Europe and the USA, was 16,152 subjects (5964 cases, 10,188 controls) from the discovery and replication stages and identified the *CLU* (also known as *APOJ*) and *PICALM* loci [60]. The second study also



**Fig. 3** Timeline of the discovery of AD genes/loci. Red color labelled *APP*, *PSEN1*, and *PSEN2* genes are for early-onset AD. Green color labelled *NFIC* and *OR2B2* loci were discovered in transeth-

nic GWAS. Purple color labelled *IGF1R* locus is unique to African Americans and blue color labelled *SCARB2* locus is unique to Japanese. The remaining loci were discovered in Europeans/Whites

identified *CLU* as well as *CR1* in a total discovery and replication sample of 14,636 subjects (6010 cases, 8625 controls) from the European Alzheimer's Disease Initiative (EADI) [61]. Another novel locus, *BIN1*, was identified in 2010 in a three-stage analysis of new and previously published GWAS on 35,336 subjects (8371 cases, 26,965 controls) derived from the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium, Translational Genomics Research Institute (TGEN), Mayo AD GWAS, EADI, and GERAD consortium [62]. In 2011, five new loci (*CD2AP*, *EPHA1*, *MS4A4A/MS4A6A*, *ABCA7*, and *CD33*) were reported in two back-to-back publications in *Nature Genetics* [63, 64]. The first study comprising 59,716 individuals (19,870 cases, 39,846 controls) from GERAD, GERAD 2, EADI, EADI 2, Mayo 2, CHARGE, Alzheimer's Disease Neuroimaging Initiative (ADNI), and TGEN (collectively called GERAD+) in a 3-stage study design identified the *ABCA7* and *MS4A6A/MS4A4E* loci. The second study used a 2-stage study consisting of 22,771 subjects (11,840 cases, 10,931 controls) derived from 12 cohorts and the U.S. National Institute on Aging (NIA)-funded Alzheimer Disease Centers, as part of the Alzheimer Disease Genetics Consortium (ADGC), and identified the *MS4A4/MS4A6E* and *EPHA1* loci. Furthermore, two loci (*CD33* and *CD2AP*) were identified by combining GERAD+ and ADGC results: a) GERAD+ and 19,072 ADGC subjects (9980 cases, 9090 controls), and b) ADGC and 31,658 GERAD+ samples

(6992 cases, 24,666 controls). The different lead SNPs noted at the *MS4A* (membrane-spanning 4-domains subfamily A) locus on chromosome 11 in these two studies detected the same signal due to the strong LD in this region that contains six of the 16 *MS4A* genes. Noteworthy, evidence for suggestive association with *CD33* and *EPHA1* was also reported earlier [56, 62], but they achieved GWS only using much larger sample size. The previously reported GWS *EXOC3L2* locus on chromosome 19 [62], which is within 300 kb of *APOE*, was not found to be significant after adjusting for the effect of *APOE* [64]. The number of total loci, including *APOE*, for LOAD reached to 10 in 2011.

Realizing that even much larger sample size is needed to identify additional loci, investigators from the ADGC, CHARGE, EADI, and GERAD consortia along with additional investigators from Europe and the USA joined hands and established the International Genomics of Alzheimer's Project (IGAP) comprising a total sample of 74,046 individuals (25,580 cases and 48,466 controls) [65]. The IGAP conducted a meta-GWAS analysis on 54,162 samples in stage 1 and genotyped top SNPs in additional 19,884 samples in stage 2 and reported the identification of 11 new susceptibility loci in 2013: *INPP5D*, *MEF2C*, *HLA-DRB5/DRB1*, *NME8/GP141*, *ZCWPW1/PILRA*, *PTK2B*, *CELF1/SPI1*, *SORL1*, *FERMT2*, *SLC24A4/RIN3*, and *CASS4*. In this study, one earlier reported locus (*CD33*) was not found to be GWS, but a later study confirmed this being a genuine

**Table 2** Genome-wide significant AD-associated loci

Chr	Position (bp) <sup>a</sup>	Lead SNP	Nearest gene	Major/minor allele	MAF	OR (95% CI) <sup>b</sup>	p-value
1	1,049,997	rs113020870	<i>AGRN</i>	C/T	0.004	1.07 (1.05–1.10)	3.8E–08
1	161,185,602	rs4575098 (3'UTR)	<i>ADAMTS4</i>	G/A	0.239	1.02 (1.01–1.02)	2.1E–10
1	109,345,810	rs141749679 (p.Lys165Glu)	<i>SORT1</i>	T/C	0.004	1.38 (1.24–1.54)	7.5E–09
1	207,518,704	rs6656401	<i>CRI1/CD55</i>	G/A	0.197	1.18 (1.14–1.22)	5.7E–24
2	9,558,882	rs72777026	<i>ADAM17</i>	A/G	0.144	1.06 (1.04–1.08)	2.7E–08
2	37,304,796	rs17020490	<i>PRKD3</i>	T/C	0.145	1.06 (1.04–1.08)	3.3E–09
2	65,381,229	rs268134	<i>SPRED2</i>	A/G	0.250	0.94 (0.92–0.96)	1.5E–08
2	105,749,599	rs143080277	<i>NCK2</i>	T/C	0.004	1.45 (1.34–1.63)	1.3E–12
2	127,135,234	rs6733839	<i>BINI</i>	C/T	0.407	1.20 (1.17–1.23)	2.1E–44
2	202,878,716	rs139643391 (3'UTR)	<i>WDR12</i>	TC/T	0.131	0.94 (0.92–0.96)	1.1E–08
2	233,117,202	rs10933431	<i>INPP5D</i>	C/G	0.223	0.91 (0.88–0.94)	3.4E–09
3	57,192,122	rs184384746	<i>HESX1</i>	C/T	0.002	1.21 (1.14–1.30)	1.2E–08
3	155,069,722	rs16824536	<i>MME</i>	G/A	0.054	0.92 (0.89–0.95)	3.6E–08
	155,084,189	rs61762319 (Met8Val)	<i>MME</i>	A/G	0.026	1.16 (1.11–1.21)	2.2E–11
4	993,555	rs3822030	<i>IDUA</i>	T/G	0.429	0.95 (0.94–0.96)	8.3E–12
4	11,025,995	rs4351014	<i>CLNK</i>	T/C	0.265	0.93 (0.91–0.95)	2.6E–11
4	40,197,226	rs2245466 (5'UTR)	<i>RHOH</i>	C/G	0.343	1.05 (1.03–1.06)	1.2E–09
4	76,217,307	rs920608	<i>SCARB2/FAM47E</i>	A/C	0.038	0.65 (0.57–0.75)	5.3E–09 <sup>e</sup>
5	14,724,304	rs112403360	<i>ANKH</i>	T/A	0.073	1.09 (1.06–1.12)	2.3E–09
5	86,927,378	rs62374257	<i>COX7C</i>	T/C	0.23	1.07 (1.05–1.09)	1.4E–15
5	88,927,603	rs190982	<i>MEF2C</i>	A/G	0.408	0.92 (0.89–0.95)	3.2E–08
5	140,335,105	rs2074612	<i>APBB3/HBEGF</i>	C/T	0.438	1.08 (1.05–1.11)	8.0E–09
5	151,052,827	rs871269	<i>TNIP1</i>	C/T	0.32	0.98 (0.97–0.99)	1.4E–09
5	157,099,320	rs6891966	<i>HAVCR2</i>	G/A	0.23	0.98 (0.97–0.99)	7.9E–10
5	180,201,150	rs113706587	<i>RASGEF1C</i>	G/A	0.11	1.09 (1.07–1.12)	2.2E–16
6	27,915,491	rs1497525	<i>OR2B2</i>	C/A	0.05–0.08	1.34 (1.15–1.56)	2.1E–08 <sup>d</sup>
6	32,592,048	rs34855541	<i>HLA-DRB1/DRB5</i>	A/G	0.135	0.90 (0.87–0.92)	9.5E–15
6	41,161,514	rs75932628 (p.Arg47His)	<i>TREM2</i>	C/T	0.008	2.01 (1.65–2.44)	2.7E–15
6	47,520,026	rs10948363	<i>CD2AP</i>	A/G	0.266	1.10 (1.07–1.13)	5.2E–11
6	114,291,731	rs785129	<i>HS3ST5</i>	C/T	0.35	1.04 (1.03–1.06)	2.4E–09
7	7,817,263	rs6943429	<i>UMAD1</i>	C/T	0.42	1.05 (1.03–1.06)	1.0E–10
7	8,204,382	rs10952097	<i>ICA1</i>	C/T	0.114	1.07 (1.05–1.10)	6.8E–09
7	12,229,132	rs5011436	<i>TMEM106B</i>	A/C	0.41	1.02 (1.01–1.02)	2.7E–09
7	28,129,127–28,129,134	rs1160871	<i>JAZF1</i>	GTCTT/G	0.222	0.95 (0.93–0.97)	9.8E–09
7	37,801,932	rs2718058	<i>GPR141/NME8</i>	A/G	0.373	0.93 (0.90–0.95)	4.8E–09
7	54,873,635	rs76928645	<i>SEC61G</i>	C/T	0.103	0.93 (0.91–0.95)	1.6E–10
7	100,374,211	rs1859788 (p.Gly78Arg)	<i>PILRA</i>	G/A	0.321	0.98 (0.98–0.99)	3.3E–18
7	143,402,040	rs10808026	<i>EPHA1/ZYX</i>	C/A	0.199	0.90 (0.88–0.93)	1.3E–10
7	146,252,937	rs114360492	<i>CNTNAP2</i>	C/T	<0.001	1.19 (1.12–1.26)	2.1E–09
8	11,844,613	rs1065712 (3'UTR)	<i>CTSB</i>	G/C	0.053	1.09 (1.06–1.12)	1.9E–09
8	27,362,470	rs73223431	<i>PTK2B</i>	C/T	0.367	1.10 (1.07–1.13)	6.3E–14
8	27,610,169	rs9331896	<i>CLU</i>	T/C	0.387	0.88 (0.85–0.90)	4.6E–24
8	144,103,704	rs34173062	<i>SHARPIN</i>	G/A	0.081	1.13 (1.09–1.16)	1.7E–18
9	104,903,697	rs1800978	<i>ABCA1</i>	C/G	0.13	1.06 (1.04–1.08)	1.6E–09
10	11,678,309	rs7920721	<i>ECHDC3</i>	A/G	0.389	1.08 (1.05–1.11)	2.3E–09
10	60,025,170	rs7068231	<i>ANK3</i>	G/T	0.403	0.95 (0.94–0.96)	3.3E–13
10	59,886,075	rs1171814	<i>CCDC6</i>	G/T	0.481	0.95 (0.93–0.97)	3.8E–08
10	80,520,381	rs1878036	<i>TSPAN14</i>	T/G	0.207	1.07 (1.05–1.10)	2.7E–09
10	96,266,650	rs6584063	<i>BLNK</i>	A/G	0.043	0.89 (0.86–0.92)	6.7E–11
10	122,413,396	rs7908662	<i>PLEKHA1</i>	A/G	0.467	0.96 (0.95–0.97)	2.6E–09
11	47,358,789	rs3740688	<i>SPI1/CELF1</i>	T/G	0.448	0.92 (0.89–0.94)	5.4E–13
11	60,169,453	rs7933202	<i>MS4A2</i>	A/C	0.391	0.89 (0.87–0.92)	1.9E–19
11	86,156,833	rs10792832	<i>PICALM</i>	G/A	0.358	0.87 (0.85–0.89)	9.3E–26
11	121,564,878	rs11218343	<i>SORL1</i>	T/C	0.040	0.80 (0.75–0.85)	2.9E–12

**Table 2** (continued)

Chr	Position (bp) <sup>a</sup>	Lead SNP	Nearest gene	Major/minor allele	MAF	OR (95% CI) <sup>b</sup>	p-value
12	113,281,983	rs6489896	<i>TPCN1</i>	T/C	0.076	1.08 (1.05–1.10)	1.8E–09
14	52,924,962	rs17125924	<i>FERMT2</i>	A/G	0.099	1.12 (1.08–1.15)	1.3E–11
14	92,460,608	rs10498633	<i>RIN3/SLC24A4</i>	G/T	0.217	0.91 (0.88–0.94)	5.5E–09
14	105,761,758	rs7157106	<i>IGH</i> cluster	G/A	0.36	1.05 (1.03–1.07)	2.0E–08
	106,665,591	rs10131280	<i>IGH</i> cluster	G/A	0.133	0.94 (0.92–0.96)	4.3E–10
15	50,709,337	rs59685680	<i>SPPL2A</i>	T/G	0.198	0.92 (0.89–0.95)	7.3E–09
15	58,753,575	rs593742	<i>ADAM10</i>	A/G	0.298	0.93 (0.91–0.95)	2.8E–11
15	63,277,703	rs117618017 (p.Thr27Ile)	<i>APH1B</i>	C/T	0.139	1.09 (1.06–1.12)	1.5E–08
15	64,131,307	rs3848143	<i>SNX1</i>	A/G	0.22	1.05 (1.04–1.07)	8.4E–11
15	64,433,291	rs74615166	<i>TRIP4</i>	T/C	0.02	1.31 (1.19–1.44)	9.7E–09
15	78,936,857	rs12592898	<i>CTSH</i>	G/A	0.133	0.94 (0.92–0.96)	4.2E–09
15	97,449,455	rs570487962	<i>IGF1R</i>	Rare variants	0.01	0.10 (0.05–0.22)	1.6E–09 <sup>c</sup>
16	19,796,841	rs7185636	<i>IQCK</i>	T/C	0.156	0.92 (0.89–0.95)	2.4E–08
16	30,010,081	rs1140239	<i>DOC2A</i>	C/T	0.379	0.94 (0.93–0.96)	2.6E–13
16	31,111,250	rs889555	<i>KAT8</i>	C/T	0.29	0.95 (0.94–0.97)	3.2E–08
16	70,660,097	rs4985556 (p.Tyr213Ter)	<i>IL34</i>	C/A	0.088	1.09 (1.05–1.12)	3.7E–08
16	79,321,960	rs62039712	<i>WWOX</i>	G/A	0.094	1.16 (1.09–1.24)	3.7E–08
16	79,574,511	rs450674	<i>MAF</i>	T/C	0.373	0.96 (0.95–0.98)	3.2E–08
16	81,908,423	rs72824905 (p.Pro522Arg)	<i>PLCG2</i>	C/G	0.009	0.68 (0.60–0.77)	5.4E–10
16	86,420,604	rs16941239	<i>FOXF1</i>	T/A	0.029	1.13 (1.08–1.17)	1.3E–08
16	90,103,687	rs56407236	<i>PRDM7</i>	G/A	0.069	1.11 (1.08–1.14)	6.5E–15
17	1,728,047	rs35048651 (5'UTR)	<i>WDR81</i>	TGAG/T	0.214	1.06 (1.04–1.08)	7.7E–11
17	5,233,752	rs7225151	<i>RABEP1/SCIMP</i>	G/A	0.118	1.10 (1.07–1.13)	6.1E–12
17	18,156,140	rs2242595	<i>MYO15A</i>	G/A	0.112	0.94 (0.92–0.96)	1.1E–09
17	44,364,976	rs708382	<i>GRN</i>	T/C	0.39	1.02 (1.01–1.02)	2.0E–09
17	46,275,856	rs2732703	<i>KANSL1/LRRC37A</i>	T/G	0.13	0.73 (0.65–0.81)	5.8E–09 <sup>f</sup>
17	49,219,935	rs616338 (p.Ser209Phe)	<i>ABI3</i>	C/T	0.008	1.43 (1.28–1.60)	4.6E–10
17	58,320,645	rs2526380	<i>BZRAP1</i>	G/C	0.44	0.97 (0.96–0.98)	2.6E–08
	58,331,728	rs2632516	<i>TSPOAP1-ASI</i>	G/C	0.44–0.60	0.92 (0.91–0.94)	4.4E–08 <sup>g</sup>
17	63,460,787	rs138190086	<i>ACE/CYB561</i>	G/A	0.017	1.25 (1.16–1.35)	1.9E–09
18	58,522,227	rs76726049	<i>ALPK2</i>	T/C	0.011	1.06 (1.04–1.08)	3.3E–08
19	1,056,493	rs3752246	<i>ABCA7</i>	C/G	0.182	1.15 (1.11–1.18)	3.1E–16
19	1,854,255	rs149080927 (3'UTR)	<i>KLF16</i>	G/C/G	0.48	1.05 (1.04–1.07)	5.1E–10
19	3,405,594	rs9749589	<i>NFIC</i>	T/A	0.02–0.16	0.76 (0.69–0.83)	1.5E–08 <sup>h</sup>
19	44,908,684	rs429358 (p.Cys112Arg)	<i>APOE</i>	T/C	0.216	3.32 (3.20–3.45)	1.2E–881
19	45,738,583	rs76320948	<i>AC074212.3</i>	C/T	0.046	0.97 (0.96–0.98)	4.6E–08
19	48,710,247	rs2452170	<i>NTN5</i>	A/G	0.47	0.99 (0.99–1.00)	1.7E–08
19	49,950,060	rs9304690	<i>SIGLEC11</i>	C/T	0.24	1.05 (1.03–1.07)	4.7E–09
19	51,224,706	rs3865444 (5'UTR)	<i>CD33</i>	C/A	0.336	0.99 (0.98–0.99)	6.3E–09
19	54,313,903	rs1761461	<i>LILRB2</i>	A/C	0.49	1.01 (1.01–1.02)	1.6E–09
20	413,334	rs1358782	<i>RBCK1</i>	G/A	0.246	0.95 (0.94–0.97)	1.6E–08
20	56,423,488	rs6014724	<i>CASS4</i>	A/G	0.088	0.89 (0.87–0.93)	1.1E–10
20	63,743,088	rs6742	<i>SLC2A4RG</i>	C/T	0.221	0.95 (0.93–0.97)	2.6E–09
21	26,784,537	rs2830500	<i>APP/ADAMTS1</i>	C/A	0.336	0.93 (0.91–0.95)	2.6E–08

Chr = chromosome, MAF = minor allele frequency and the effect allele

<sup>a</sup>GRCh38 assembly

<sup>b</sup>Odds ratio with 95% confidence interval

<sup>c</sup>Among Japanese

<sup>d</sup>Transethnic GWAS including Japanese and Europeans

<sup>e</sup>Among African Americans

<sup>f</sup>Among non-*APOE*\*4

<sup>g</sup>Transethnic GWAS in which MAF varied from 0.44 to 0.60 in different ethnic groups

<sup>h</sup>Transethnic GWAS in which MAF varied from 0.02 to 0.16 in different ethnic groups, *p*-value represents interaction term (*NFIC*/rs9749589×*APOE*\*4)



locus (discussed below). The potential of *PTK2B* being an AD locus was also reported in an earlier GWAS with a small sample size ( $n=2229$ ) in which multiple SNPs with  $p < 4E-03$  were reported in this gene [66]. A rare GWS variant in *TREM2* (p.Arg47His/rs75932628) was also discovered in 2013 [67] and subsequently replicated in independent studies [68]. In 2014, a follow-up meta-analysis of the summary statistics of suggestive IGAP loci with a Spanish sample of 4372 subjects (1808 case, 2564 controls) resulted in the identification of a novel GWS signal with a rare intronic *TRIP4* variant [69]. The IGAP data was further used in 2017 to identify rare coding variants in a 3-stage study using a whole-exome microarray in 34,174 samples in stage 1 followed by genotyping 35,962 independent samples in stage 2 and 14,997 samples in stage 3, totaling 85,133 subjects (37,022 cases, 48,402 controls) [70]. This study confirmed an earlier reported *TREM2*/p.Arg47His variant and identified another GWS *TREM2* variant (p.Arg63His) whose effect was independent from p.Arg47His. This study also reported GWS rare missense coding variants in novel *PLCG2* (p.Pro522Arg) and *ABI3* (p.Ser209Phe) genes.

In addition to the standard case–control genetic association study in GWAS that includes a set of subjects with a defined disease (cases) and a set of subjects who do not have a disease (controls), an alternative GWAS approach is a proxy case–control genetic association study in which relatives of affected (proxy cases) and un-affected (proxy controls) individuals are identified from a population-based cohort study. This approach is called as a genome-wide association study by proxy (GWAX) [71]. Although GWAX has lower power to detect association using proxy cases, it increases the sample size substantially and logistically it is more feasible than collecting standard cases. This approach has shown to be a valid approximation of future disease status for 12 common diseases, including AD, by showing replication of established genetic associations in GWAX. During 2017–2019, three AD GWAX were published using the UK Biobank (UKB) cohort, which were meta-analyzed with data from standard GWAS [71–73]. The first GWAX done in 2017 [71] used a subset of the UKB cohort comprising 14,482 proxy cases and 100,082 proxy controls (total 114,564 subjects) and identified *APOE\*4*/rs429358 as the only GWS signal ( $p=9.7E-195$ ). Next, GWAX data was meta-analyzed with the IGAP-2013 discovery GWAS data ( $n=74,046$ ), making the combined sample of 188,610 subjects (40,062 cases, 148,548 controls), which resulted in the identification of 4 novel loci: *HBEGF/APBB3*, *ECHDC3*, *SPPL2A*, and *SCIMP/RABEP1*. Interestingly, although all 4 loci were suggestively significant in the IGAP data ( $p$  range =  $5.01E-07$  to  $1.63E-07$ ), they were only nominal significant in the GWAX data ( $p$  range =  $0.04$  to  $0.00026$  to). Of the 26 top GWS SNPs, including *APOE\*4*, 9 did not even achieve nominal significance in the GWAX data

( $p$  range =  $0.46$  to  $0.059$ ), indicating low power with proxy cases. Furthermore, *HBEGF* and *ECHDC3* loci were also identified in a transethnic GWAS published in 2017 (discussed below). The second GWAX/GWAS published in 2018 [72] included 314,278 individuals (27,696 maternal proxy cases, 14,338 paternal proxy cases) from the UKB cohort, also including subjects from GWAX-2017. GWAX was performed separately for maternal and paternal AD due to a 1.7-fold difference in disease prevalence — 9.6% and 5.5%, respectively. Six known GWS loci were identified in the UKB paternal and maternal GWAS meta-analysis. Then, UKB parental summary statistics were meta-analyzed with the IGAP-2013 discovery GWAS data (combined sample of 388,324 subjects, including 67,614 cases and 320,710 controls), which resulted in the identification 27 loci including 4 new ones: *ADAM10*, *KAT8*, *IL34*, and *ACE*. The third GWAX/GWAS published in 2019 [73] further expanded the sample to 455,258 subjects (71,880 cases, 383,378 controls) by meta-analysis in a 3-stage study. Stage 1 involved meta-analysis of clinically diagnosed 79,145 subjects (24,087 cases, 55,058 controls) derived from IGAP-2013 ( $n=54,162$ ), Alzheimer's disease working group of the Psychiatric Genomics Consortium (PGC-ALZ;  $n=17,477$ ), and whole-exome sequencing (WES) data from the Alzheimer's Disease Sequencing Project (ADSP;  $n=7,506$ ), which identified 18 known GWS loci. Stage 2 performed GWAX analysis on 376,133 individuals (47,793 proxy cases, 328,320 proxy controls) and identified 13 GWS loci, 8 of which overlapped with stage 1 and the lead SNPs of 3 other loci were not available in the stage 1 sample, including one in *TREM2*; one was in an earlier GWAX/GWAS-2018 [72]. Due to a strong genetic correlation (0.81) observed between clinically assessed AD status and AD-by-proxy as well as high concordance in the individual SNP effects in stage 1 and 2 analyses, meta-analysis of stage 1 and 2 results was conducted in stage 3 on the combined sample of 455,258. The combined meta-analysis identified 29 distinct loci, of which 7 were new: *ADAMTS4*, *HESX1*, *CLNK*, *CNTNAP2*, *APHIB*, *ALPK2*, and *AC074212.3*. The lead SNPs at 7 of the novel 9 loci were further tested in an independent Icelandic cohort (deCODE), including 6593 AD cases and 174,289 controls and only 5 of them were replicated at nominal significance. Four loci identified in IGAP-2013 (*MEF2C*, *NME8*, *CELF1*, and *FERMT2*) were not found to be GWS in this analysis, mostly due to the lower association signals in the UKB data set. However, the *CD33* locus, which was not GWS in IGAP-2013, was significant in the combined meta-analysis. *MEF2C* was also found to be GWS in the 2017 GWAX/GWAS [71].

In 2019, IGAP also expanded its stage 1 sample to 63,926 and along with stage 2 ( $n=18,845$ ) and stage 3 ( $n=11,666$ ) samples performed a meta-analysis on clinically assessed sample of 94,437 individuals (35,272 cases and 59,163

controls) [74]. This paper was published back-to-back with the GWAX/GWAS-2019 paper [73] and identified 25 GWS loci, including 3 new ones: *IQCK*, *WVOX*, and *ADAMTS1* [74]. IGAP-2019 also identified GWS *ADAM10* and *ACE* loci, which overlapped with GWAX/GWAS findings [72, 73]. It is noteworthy that the larger GWAX/GWAS-2019 dataset [73] showed no or only modest evidence of association for *IQCK* ( $p=0.14$ ), *ADAMTS1* ( $p=0.016$ ), and *WVOX* ( $p=0.77$ ). Two loci that were GWS in IGAP-2013 did not achieve GWS threshold in IGAP-2019 (*NME8*:  $p=2.7E-07$  and *MEF2C*:  $p=9.1E-08$ ).

More recently, two new GWAX/GWAS have been published using the largest discovery sample for family history of AD to date in the UKB. The first study included 53,042 unique individuals who were either diagnosed with AD or who reported a parent or sibling having dementia, and 355,900 controls, totaling 408,942 subjects [75]. This analysis identified 13 GWS loci, of which 3 were new (*NCK2*, *PRL*, and *FAM135B*). Next, GWAX results were meta-analyzed with the IGAP-2019 GWAS stage 1 sample of 63,926, resulting in a total sample of 472,868 subjects (75,024 cases and 397,844 controls). This identified 34 AD risk loci, including 4 novel loci near *SPRED2*, *NCK2*, *CCDC6*, and *TSPAN14*. Two additional novel loci achieved subthreshold GWS: *TMEM163* on chromosome 2 ( $p=5.24E-08$ ) and *IKZF1* on chromosome 7 ( $p=7.68E-08$ ). The *PRL* and *FAM135B* loci identified in the GWAX discovery sample showed no evidence of association in the IGAP-2019 stage 1 sample. Replication of the identified 34 loci was further sought in two data sets, including 7409 subjects from the GR@ACE study (4120 cases and 3289 controls) and 135,638 subjects from the FinnGen Biobank v.3 (3697 cases and 131,941 controls). Only about half of the 34 loci were nominally replicated ( $p < 0.05$ ) in either of the two replication samples. Subsequently, meta-analysis was performed with all four datasets comprising a combined sample of 615,915 (82,841 cases, 533,074 controls), in which support for 3 of the 4 new loci (*NCK2*, *CCDC6* and *TSPAN14*) was increased, but weakened for *SPRED2* ( $p=1.3E-07$ ). The second GWAX/GWAS expanded the study subjects to 1,126,563 that included 90,338 (46,613 proxy) AD cases and 1,036,225 (318,246 proxy) controls and reported 38 loci [76], including 7 novel loci (*AGRN*, *TNIP1*, *HAVCR2*, *TMEM106B*, *GRN*, *NTN5*, *LILRB2*). Two of the novel AD loci (*TMEM106B*, *GRN*) are also known frontotemporal dementia (FTD) genes [77, 78], suggesting their potential roles in protein clearance rather than in specific disease-related protein aggregates [76]. In a recent paper available in preprint [79] using the largest number of 111,326 (46,828 proxy) AD cases along with 677,633 controls have identified the most number of 68 loci in one study that also included 35 new loci: *SORT1*, *ADAM17*, *PRKD3*, *WDR12*, *MME*, *IDUA*, *RHOH*, *ANKH*, *COX7C*,

*RASGEF1C*, *HS3ST5*, *UMAD1*, *ICA1*, *JAZF1*, *SEC61G*, *CTSB*, *SHARPIN*, *ABCA1*, *ANK3*, *BLNK*, *PLEKHA1*, *TPCN1*, *IGH* cluster, *SNX1*, *CTSH*, *DOC2A*, *MAF*, *FOXF1*, *PRDM7*, *WDR81*, *MYO15A*, *KLF16*, *SIGLEC11*, *RBCK1*, and *SLC2A4RG*. Another paper in preprint [80] has identified two of these loci (*SHARPIN*, *ATF/SIGLEC11*) in a total sample of 80,685 AD cases and 243,682 controls.

### APOE-Stratified GWAS

*APOE\*4* is the greatest risk factor for AD and is also the major contributor to the observed genetic variance. It is possible that in the presence of *APOE\*4*, the effect of some significant genes may not be apparent or additional genetic factors may act in concert with *APOE\*4* to increase AD risk. In order to address this question, the first *APOE*-stratified GWAS-based on the presence (*E4* carriers) or absence (non-*E4* carriers) of *APOE\*4* was carried out in 2016 [81] on the IGAP-2013 sample comprising 19,559 *APOE\*4* carriers (10,352 cases, 9207 controls) and 34,152 non-*APOE\*4* carriers (7184 cases, 26,968 controls). Suggestive associations at  $p < 1E-04$  were examined in an independent sample comprising 1786 *E\*4* carriers (1250 cases, 536 controls) and 2417 non-*E\*4* carriers (718 cases, 1699 controls). A GWS signal was observed among non-*APOE\*4* carriers on chromosome 17 between *KANSL1* and *LRRC37A*, which is about 200 kb downstream from the *MAPT* gene that codes for the microtubule-associated protein tau found in AD neurofibrillary tangles. However, conditional analysis excluded the involvement of the *MAPT* gene or another gene distal to *LRRC37A*. Two subsequent WES studies using the overlapping ADSP dataset nominated *NSF* as a candidate gene in this region, which is located downstream from the *LRRC37A* gene [82, 83]. *NSF/rs199533* is a coding variant (p.Lys702Lys), which was found to be the most significant *cis*-eQTL for *KANSL1-AS1* and *LRRC37A* in different brain regions and blood [83]. The *APOE*-stratified analysis also revealed that the previously established associations of *MSA4* and *TREM2* are mainly derived from non-*APOE\*4* carriers.

Altogether, 90 new loci were identified in addition to *APOE*, among White or European populations during the period from 2009 to early 2022 in at least one GWAS (Table 2). Although the combined sample size in AD GWAS and/or GWAX/GWAS seems large, there is considerable overlap in samples used in these studies and thus technically these studies are not independent.

### GWAS in Non-Whites

Although more older Whites are living with AD, older Black/African Americans and Hispanics are disproportionately more likely than older Whites to have AD [2]. Despite this fact, these population groups along with other

non-White populations are understudied with regards to the genetic susceptibility to AD. The first large GWAS on African Americans age 60 and above was done in 2013 under the banner of ADGC, which comprised 5896 subjects (1968 case, 3928 control) [84]. In addition to *APOE*, the *ABCA7* locus was found to be GWS. Interestingly, although *APOE*\*4 allele in an African genetic background confers lower risk than that observed for Whites, the effect size of *ABCA7* was double in African Americans (OR = 2.3) than reported previously in Whites. An expanded ADGC African American cohort, the largest to date, including 8006 subjects (2784 cases and 5222 controls), has been published recently [85], in which *APOE* was the only GWS locus and it replicated *ABCA7*/rs115550680 and 5q35/rs145848414 variants at  $p < 5E - 06$  from the 2013-GWAS. Five additional GWS loci observed in Whites (*BINI*, *TREM2*, *CD2AP*, *FERMT2*, and *WVVOX*) were also replicated with nominal significance.

After adjusting for the effect of *APOE*, a novel GWS intergenic locus associated with rare variants was observed in African Americans on chromosome 15q26 near *ARRDC4* and *IGF1R* (Table 2). Another locus associated with rare variants on chromosome 11 was subthreshold GWS (*API5*/rs569584007; MAF = 0.01;  $p = 8.8E - 08$ ). Five additional novel loci with rare variants (*SIPA1L2*, *WDR70*, *ACER3*, *PIK3C2G*, *RBFOX1*) and 4 with common variants (*EDEM1*, *ALCAM*, *GPC6*, *VRK3*) were observed at  $p < 1E - 06$ .

The previously reported AD-associated rare *COBL*, *AKAP9*, and *TREM2* variants in African Americans [86–89] were replicated in this larger GWAS as follows: *COBL*/rs112404845 (200 kb upstream of *COBL*) at  $p = 5.4E - 06$ , *AKAP9*/rs149979685 (p.Ser3771Leu) at  $p = 0.005$ , *AKAP9*/rs914662445 (p.Ile2558Met) at  $p = 0.01$ , *TREM2*/rs2234256 (p.Leu211Pro) at  $p = 0.001$ , *TREM2*/rs2234258 (p.Trp191X) at  $p = 1.4E - 03$ , and *TREM2*/rs7748513 at  $p = 3.6E - 05$ . The latter *TREM2* variant is 1 kb downstream of and in LD with the strongest associated *TREM2*/rs75932628 (p.Arg47His) variant in Whites, which is present sporadically in African Americans (MAF = 0.0009) [88], perhaps due to white admixture. The rare *COBL* and *AKAP9* variants are unique to people with African ancestry. Despite the fact that this was the largest GWAS in African Americans, its sample size compared with White GWAS is extremely small and thus only one GWS and some suggestive novel loci were identified.

A Chinese study that included 447 AD cases and 442 cognitively normal controls and 1745 mixed controls from the CONVERGE (China Oxford and Virginia Commonwealth University Experimental Research on Genetic Epidemiology) found *APOE* as the only GWS locus [90]. This study also tested 16 of the 21 IGAP-2013 GWS variants and found only 3 of them (*BINI*, *CD2AP*, *FERMT2*) to be nominal significant in Chinese. An earlier study replicated *PICALM*, *BINI*, *CLU*, and *MS4A4E* variants in Chinese [91]. Although

known missense risk variants in *TREM2* have not been replicated in Chinese, a rare *TREM2*/rs2234255 coding variant in exon 3 (p.His157Try) has shown to confer a considerable risk for AD in Chinese [92].

Recently, a large two-stage GWAS conducted in Japanese comprising 5178 cases and 6520 controls identified a novel *SCARB2/FAM47E* locus on chromosome 4 in addition to the known *APOE* and *SORL1* loci [93]. The lead SNP and its proxies were eQTLs for both *FAM47E* and *SCARB2* genes in several brain tissues. Previously, the *SCARB2/FAM47E* locus has also been implicated with Parkinson disease (PD) [94]. *SCARB2* seems to be the candidate gene for both AD and PD as it plays a role in neuroinflammation and in the degradation process of  $\alpha$ -synuclein [95].

### Transethnic GWAS

The transethnic meta-analysis approach that gathers results from multiethnic participants in existing studies has been found to be useful in discovering new loci for genetic traits and diseases [96–99]. This approach has enabled the identification of multiple associations previously unreported in European-only analyses [99]. In an effort to discover new AD loci, a two-stage transethnic GWAS was conducted in 2017 [100]. The stage 1 sample comprised 33,263 subjects, including 26,320 American Whites, 4983 African Americans, 1845 Japanese, and 115 Israeli Arabs as part of the ADGC. Stage 2 sample consisted of 26,287 White subjects (5813 cases, 20,474 controls) from IGAP-2013 after excluding the ADGC datasets. GWAS meta-analysis from both stages identified 3 GWS loci near *PFDNI/HBEGF* on chromosome 5, *USP6NL/ECHDC3* on chromosome 10, and *TSPOAP1* (formally *BZRAP1*)/*TSPOAP1-ASI* on chromosome 17. A fourth locus was identified based on a GWS interaction between *APOE*\*4 and *NFIC*/rs9749589 SNP on chromosome 19 (Table 2) in which this SNP was protective in *APOE*\*4 carriers (OR = 0.86; 95% CI: 0.80–0.83;  $p = 5.5E - 05$ ), but with slight increased risk in non-*APOE*\*4 carriers (OR = 1.12; 95% CI: 1.05–1.18;  $p = 1.2E - 03$ ). Further evaluation of six GWS intergenic SNPs between *USP6NL* and *ECHDC3* genes revealed that these associations were exclusive to non-*APOE*\*4 subjects (leading SNP: rs7920721,  $p = 2.7E - 09$ ) [100]. As described above, two of the 4 loci identified in the transethnic GWAS (*PFDNI/HBEGF*, *USP6NL/ECHDC3*) produced suggestive evidence of association ( $p > E - 07$ ) in IGAP-2013 and then became GWS after combining the IGAP-2013 data with 2017 GWAS. Thus, in 2017, these two loci were identified independently by this transethnic GWAS as well as GWAS/GWAS using the IGAP-2013 data. It is noteworthy that although the discovery sample in the transethnic GWAS was much smaller than used in the recent large Whites GWAS and GWAS, it still enabled the identification of novel loci.



This highlights the value of combining data from diverse population groups in which diversity can be used to increase power for gene discovery [101]. Considering that *PFDN1/HBEGF* and *USP6NL/ECHDC3* loci were also identified in GWAX/GWAS-2017, *TSPOAPI1/TSPOAPI-ASI* and *NFIC* loci can exclusively be attributed to the transethnic GWAS. Interestingly, the transethnic *TSPOAPI1/TSPOAPI-ASI*/rs2632516 SNP was also GWS ( $p = 1.4E - 09$ ) in the large case-control phase 1 sample of > 79 k, but it was less significant in the phase 2 proxy sample ( $p = 0.005$ ) that yielded a meta- $p$  of  $9.7E - 07$  [73]. However, another close by SNP in this region, rs2526380, was GWS in the meta-analysis ( $p = 2.6E - 08$ ) in the 2019-GWAX/GWAS [73]. Thus, *TSPOAPI1/TSPOAPI-ASI* locus was technically GWS significant in 2019-GWAX/GWAS, but it was not highlighted in the paper as such. Due to this reason, both the transethnic and GWAX/GWAS SNPs are included in Table 2.

A recent transethnic GWAS meta-analysis, comprising Japanese (5178 cases and 6520 controls) and Whites from the 2019 IGAP stage 1 data (21,982 AD cases and 41,944 controls), identified a novel locus, *OR2B2*, on chromosome 6 [93]. Future transethnic GWAS including larger non-white populations will help to identify additional AD novel loci.

In summary, GWAS and GWAX/GWAS have identified 95 loci for LOAD, which were GWS in at least one GWAS (Table 2). Of the 95 loci, one is unique to African Americans (*IGFIR*), one is unique to Japanese (*SCARB2/FAM47E*), and two were discovered using the transethnic GWAS approach (*NFIC*, *OR2B2*). Of the remaining 91 loci in Whites, one was discovered in the *APOE*-stratified analysis (*KANSL1*). Although the rare coding *COBL* and *AKAP9* variants observed in African American GWAS were not GWS, it seems that the *COBL* and *AKAP9* are genuine candidate genes for LOAD in people with African ancestry.

## Sequencing Strategy to Identify Additional AD Variants/Genes

As discussed above, substantial progress has been made via large-scale GWAS as well as meta-analyses of GWAS that have identified about 95 susceptibility loci for LOAD. However, GWAS-implicated genes/variants explain only a portion of the AD genetic variance [9–11] and much of the genetic variance remains unexplained. GWAS arrays that usually rely on LD to detect association signals fail to detect phenotypic association with functional or causal variants that are neither genotyped nor in LD with SNPs included in GWAS arrays even using imputation methods [102–104]. It is estimated that about 1% of the SNPs in the HapMap, which number in tens of thousands, are not in LD with other SNPs, and thus, they must directly be genotyped or sequenced for association analyses [105, 106]. To identify such variants as well as

novel genetic variation affecting AD risk, the Alzheimer's Disease Sequencing Project (ADSP) was implemented by the National Institute on Aging (NIA), and the initial phase included WES on about 11,000 AD cases and controls [103, 107]. The latest WGS data released from the ADSP in March 2021, and available to qualified investigator, was on about 15,000 AD cases and controls (<https://dss.niagads.org/niagads-dss-releases-additional-17k-whole-genomes/>) and the plan is to have WGS data from ~70,000 ethnically diverse and global cohorts in late 2022.

This first large-scale WES from ADSP [103] confirmed associations with common and rare variants in multiple previously established AD genes, including *APOE*, *ABCA7*, *HLA-DPA1*, *MS4A6A*, *PILRA*, *SORL1*, and *TREM2*. This study also identified three novel genes: one rare exome-wide significant variant in a long non-coding RNA gene, *AC099552.4* ( $p = 1.2E - 07$ ) on chromosome 7; one common nearly exome-wide significant variant in *IGHG3* ( $p = 9.8E - 07$ ) on chromosome 14; and a gene-wide significant association with *ZNF655*, including 9 high-impact rare variants ( $p = 5E - 06$ ) on chromosome 7. A follow-up *APOE*-stratified WES analysis of ADSP and additional replication cohorts identified a near GWS association with a rare *GPAI1* exonic variant, rs138412600, on chromosome 8 ( $p = 7.81E - 08$ ) among subjects lacking *APOE\*4*. On the other hand, gene-based test of rare variants identified *IGHV3-7* on chromosome 14 ( $p = 9.75E - 16$ ) and *SLC24A3* on chromosome 20 ( $p = 2.67E - 12$ ) as possible novel genes in the ADSP discovery sample among *APOE\*4* carriers [82]. As mentioned above, a recent meta-GWAS analysis has identified GWS signals in the IGH (immunoglobulin heavy chain) gene cluster on chromosome 14 ([79]; Table 2), which confirms the WES findings about the involvement of *IGHG3* and *IGHV3-7* genes in AD risk. Human IGH seems to have inherent anti-amyloidogenic activity [108] and thus have a potential role in AD pathogenesis.

About 98% of the genome is non-coding and GWAS data indicate that >90% of the disease and trait-associated variants of small effect sizes are non-coding, and many of them are concentrated in regulatory regions [109, 110]. Thus, WGS will identify additional variation in non-coding regulatory regions that will contribute to novel risk gene discovery, especially ultra-rare variants and copy number variants with large effect sizes than seen in GWAS, and will provide new molecular mechanistic insights for AD. Overall, the WGS approach offers several advantages over WES [111]: i) WGS covers both coding and non-coding regions, including the non-coding regulatory regions where GWAS have identified most of the AD-associated variants, ii) WGS even at 30× coverage has shown to be more powerful than WES for detecting coding variants, and iii) WGS provides greater uniformity of sequencing reads, determination of insertions/deletions and copy number variations, and better cost per base than WES.

An early attempt to test the utility of WGS in AD has already begun. An association study of WGS conducted on 889 Han Chinese cases and controls along with 1745 mixed controls identified two suggestive loci (*GCHI*/rs72713460,  $p = 4.4E - 05$  and *KCNJ15*/rs928771,  $p = 3.6E - 06$ ) [90]. The lead SNPs in suggestive loci were not significant in the IGAP-2013 discovery data, indicating their potential ethnic-specific associations. A recent large WGS study was carried out in a family-based sample comprising 2247 subjects from 605 multiplex AD families followed by replication in an unrelated case–control sample of 1669 (983 cases) derived from the ADSP [112]. Although no GWS association was observed with rare sequence variants ( $MAF \leq 1\%$ ), nominal meta- $p$  were reported for 13 potential novel candidate loci, including 4 from single-variant analysis (*FNBP1L*, *SEL1L*, *LINC00298*, *C15ORF41*;  $p$  range =  $1.1E - 02$  to  $2.4E - 04$ ) and 9 from spatial-clustering (gene-based) analysis (*PRKCH*, *C2CD3*, *KIF2A*, *APC*, *LHX9*, *NALCN*, *CTNNA2*, *SYTL3*, *CLSTN2*;  $p$  range =  $1.8E - 04$  to  $8.2E - 06$ ). This clearly indicates that a much larger WGS sample is needed to discover novel GWS rare variants, and the availability of larger WGS data through the ADSP in 2022 may help to identify several new AD loci.

## Molecular Pathways

Alzheimer's disease is a complex and heterogeneous disease where multiple biological pathways and their interactions appear to converge to its pathobiology. Characterization of AD-related gene network connectivity in a given pathway and its regulation and association to AD may help to provide new insights about the underlying biological mechanisms and eventually the identification of drug targets. Gene expression profiling in autopsied brain tissues from AD cases and controls has implicated distinct cell types and biological pathways in AD pathogenesis. These studies highlight the key role of microglia as well as subpopulation of oligodendrocyte and astrocyte cells and the immune and microglial biological pathways in AD [113–115]. GWAS provides a genetic approach to identify molecular pathways. In addition to the A $\beta$  pathway, new genes identified in early GWAS suggested the involvement of immunity, lipid, and endocytosis pathways [116]. Data from recent larger GWAS in conjunction with functional genomics work also implicate the role of microglia, immune system, and protein catabolism of plaques as relevant to LOAD [76, 117, 118]. In contrast to psychiatric disorders and behavioral traits where GWAS variants are primarily present within neuronal enhancers and promoters, AD-associated variants are largely confined to microglial enhancers [117]. Protein–protein interaction (PPI) network for microglia associated AD-risk genes has been found to be highly connected and centered around ApoE,

which contrasts to smaller in scope PPI networks observed for neurons and oligodendrocytes associated AD-risk genes. Whereas the microglia AD-associated genes highlight the gene ontology terms for immune function, the gene ontology terms for A $\beta$  processing were associated with neurons, microglia, and oligodendrocytes [117].

## Moving from Genomics to Functional Genomics

Genomic localization of AD susceptibility loci is only the first step toward delineating the functional genomics of AD. The next important step is to identify the functional gene or genes among the many genes located in a given locus so that their involvement in AD pathogenesis could be further investigated. The exonic location of GWAS lead or sentinel SNPs might be considered causal because they can affect protein structure and function by altering amino acid sequence (non-synonymous variants) or can affect translation or protein stability, if located in 3'UTR, and transcription binding, if located in 5'UTR. However, >90% of the GWAS-implicated variants are non-coding with no direct effect on protein structure or function and many are located far away from the closest known gene. Generally, the gene located closest to the GWAS lead SNP in a region is considered functional, which often is not the case [119].

A general framework to identify the functional gene or genes in a given locus has been proposed [120] that involves the use of conditional analyses to determine if there is a single or multiple independent signals within a locus followed by integration of disease-associated SNPs with publically available multiomic datasets. Many non-coding GWAS-associated SNPs are shown to be eQTLs (expression quantitative trait loci), suggesting they could act through altering gene expression. Other non-coding variants may also affect disease risk by altering DNA methylation (mQTLs), DNase hypersensitivity (dsQTLs), TF binding (bQTLs), or protein levels (pQTLs). Several studies have shown that GWAS-associated non-coding variants are enriched in predicted transcriptional regulatory regions, known as “*cis*-regulatory elements,” and they appear to affect disease risk by modifying the function of regulatory elements in disease-relevant cell types, with subsequent changes in target gene expression. For example, autoimmune disease-associated variants were mapped to promoters and enhancers active in B and T cells, neurological disease risk variants were mapped to promoters and enhancers active in brain tissues, and fasting blood glucose-associated variants were mapped to regulatory elements in pancreatic islets [110]. As mentioned above, AD-associated variants are largely confined to microglial enhancers followed by other brain cells [76, 113–115, 117].



As shown in Table 2, sixteen GWAS lead SNPs may directly be functional due to their locations either in coding exons (*APOE*, *SORT1*, *MME*, *TREM2*, *PILRA*, *APH1B*, *PLCG2*, *ABI3*, *IL34*) or in 3'UTR (*ADAMTS4*, *WDR12*, *CTSB*, *KLF16*) and 5'UTR (*RHOH*, *WDR81*, *CD33*). In addition to the lead coding SNPs in the above 9 genes, GWS non-synonymous variants have also been found in the *CIR*, *PVRIG* (*PILRA* locus), *MS4A6A*, *ACE*, and *CD33* and multiple genes in the *APOE* locus, including *BCL3*, *CBL3*, *BCAM*, *TOMM40*, *APOC4-APOC2*, *NKPD1*, *EXOC3L2*, and *ENSG00000267114* [73–75]. Many of these genes harboring coding variants are strong candidate genes for AD. Multiple loss-of-function coding variants in *ABCA7* are associated with AD risk ( $p=2.2E-13$ ), strongly suggesting its direct involvement in AD pathogenesis [121]. *TREM2*, *PLCG2*, and *ABI3* genes with rare coding variants are highly expressed in microglia and strongly support the involvement of microglia-mediated innate immune response in the etiology of AD. The *COBL* and *AKAP9* are also likely candidate genes for AD as rare coding risk variants discovered in these genes are unique to people with African ancestry.

Conditional analyses done on AD-associated variants in three recent large GWAS meta-analyses [73–75] identified at least two independent signals in 10 loci (*BIN1*, *PKT2B/CLU*, *ABCA7*, *NCK2*, *EPHA1*, *ADAM10*, *ACE*, *APP/ADAMTS1*, *TREM2*, and *APOE*), suggesting that there may be at least 10 additional loci which are not included in the currently known approximately 95 loci. Although *PKT2B* and *CLU* are two distinct loci on chromosome 8, there is some evidence to suggest that these two loci might physically interact to affect the AD risk through the same biological mechanism [75]. For this reason, some studies list them as one locus. However, considering that their tissue gene expression is not the same, probably they should be treated as two distinct loci until more experimental data is available. Of the two independent signals in the *APP/ADAMTS1* locus, one seems to be in the *APP* gene, in addition to the lead *ADAMTS1*/rs2830500 signal ( $p=2.6E-08$ ) located about 51 kb upstream of *ADAMTS1*. The second most significant SNP in this locus is located in an *APP* intron (rs48170900;  $p=1.0E-07$ ) [75], and its association is independent of *ADAMTS1*/rs2830500 ( $D'=0.0002$ ;  $R^2=0.0$ ), suggesting that *APP* and *ADAMTS1* are two distinct loci on chromosome 21. Noteworthy, *APP*/rs48170900 was found to be GWS ( $p=4.8E-08$ ) when a relaxed relatedness threshold was used for proxy AD cases and controls [72]. If *APP* is confirmed to be a distinct locus then this would provide further evidence of the involvement of common variants with AD risk in addition to the association of rare *APP*/rs63750847 (p.Ala673Thr) coding variant with AD. Interestingly, there were 8 GWS independent signals in the *APOE* locus, but they were not considered in association analyses due to the very strong effect of *APOE\*4* on AD risk [73].

A limitation of conditional analysis is that it can overlook functional variants if they are in strong LD with the lead SNP.

The four recent large studies also performed functional genomics analyses by employing multiple overlapping approaches to link the AD-associated variants with potential functional genes [73–76]. The IGAP-2019 study, which identified 25 loci [74], prioritized candidate genes using five strategies: i) annotation and gene-based testing for deleterious coding, loss-of-function, and splicing variants, ii) eQTL analyses, iii) evaluation of transcriptomic expression in AD clinical traits, iv) evaluation of transcriptomic expression in AD-relevant tissues, and v) gene cluster/pathway analyses. The outcome of these strategies provided strong support for the *APOE*, *ABCA7*, *BIN1*, *TREM2*, *SORL1*, *ADAM10*, *SPII*, and *CR1* as the AD risk genes. The 2019 GWAX/GWAS that identified 29 loci [73] employed genome-wide gene-based association analysis using Multimarker Analysis of GenoMic Annotation (MAGMA) along with three gene-mapping strategies implemented in functional mapping and annotation (FUMA): positional gene mapping, eQTL gene mapping, and chromatin interaction mapping. All these four strategies implicated 16 functional genes, including *HLA-DRA*, *HLA-DRB1*, *PTK2B*, *CLU*, *MS4A3*, *SCIMP*, and *RABEP1* and 9 genes in the *APOE* locus (*IGSF23*, *PVR*, *BCAM*, *PVRL2*, *TOMM40*, *APOE*, *APOC1*, *APOC4*, and *CLPTMI*). Brain-specific eQTL and mQTLs provided further functional evidence in favor of *PVRL2*, *TOMM40*, and *APOC4*. Identification of multiple potential functional genes along with multiple independent signals unveiled in conditional analyses highlights the complex LD structure in the *APOE* locus.

Schwartzentruber et al. [75] performed functional genomics analyses on 37 loci, including 34 GWS and 3 suggestive loci (*IKZF1*, *TSPOAP1*, and *TMEM163* with  $p < 5E-07$ ). Comprehensive eQTL colocalization, annotation, fine-mapping, and network analyses were employed. Excluding *APOE*, eQTL colocalization identified 80 distinct genes at 27 loci that included, among others, *PTK2B*, *BIN1*, *PILRA*, *CD33*, *TREM2*, *FCER1G*, *TSPAN14*, *APH1B*, and *ACE*. Using three distinct fine-mapping methods, 21 variants with mean causal probability > 50% across the fine-mapping methods were identified. The top 13 GWS variants with causal probability > 50% included the following: *SPRED2*/rs268120 (56% probability), *NCK2*/rs143080277 (100% probability), *BIN1*/rs6733839 (100% probability), *INPP5D*/rs10933431 (83% probability), *PILRA*/rs1859788 (60% probability), *ECHDC3*/rs7920721 (64% probability), *SORL1*/rs11218343 (100% probability), *APH1B*/rs117618017 (90% probability), *PLCG2*/rs12444183 (69% probability), *ABCA7*/rs12151021 (71% probability), *CD33*/rs12459419 (66% probability), *CASS4*/rs6014724 (55% probability), and

*ADAMTS1/rs2830489* (72% probability). Of these potential causal variants, *BINI/rs6733839C > T*, located 20 kb upstream of *BINI*, creates a binding site for the MEF2C transcription factor that facilitates hippocampal-dependent learning and memory [122], and this SNP is an eQTL for *BINI* in primary microglia and induced pluripotent stem cell-derived macrophages (IPSDMac) [123]. This suggests that this variant affects AD risk by increasing *BINI* expression in microglia and IPSDMac due to the increased binding of MEF2C. Although *BINI* and *MEF2C* are expressed in multiple tissues, their coexpression was found only in primary microglia and IPSDMac [123]. Another independent signal in *BINI*, rs13025717, is an eQTL for *BINI* in monocytes and together with rs6733839 has been implicated with AD risk through integration of single brain cell and peripheral myeloid epigenomics [124, 125]. As mentioned above, *APH1B/rs117618017* (p.Thr27Ile) is the lead GWS variant in this locus and affects the APH1B protein structure, which is a component of  $\gamma$ -secretase. This structural change in APH1B may directly affect its function and the ensuing APP processing and AD risk. However, rs117618017 was also found to be an eQTL for *APH1B* in monocytes, neutrophils, and T cells in which the effect “T” allele was associated with higher *APH1B* expression, which may also explain its association with higher AD risk [75]. *PILRA/rs1859788* (p.Gly78Arg) is also the lead GWS SNP in this locus. Although *CD33/rs12459419* (p.Ala14Val) is not the lead SNP in this locus, it is GWS ( $p = 1.3E - 08$ ) and is a strong splicing QTL, suggesting that it is likely the functional variant. Likewise, *CASS4/rs6014724* is the lead GWS SNP in this locus and is an eQTL for *CASS4* in monocytes and whole blood. *CASS4/rs17462136* is another GWS variant ( $p = 1.01E - 09$ ), which is located in 5'UTR and is predicted to decrease transcription binding [75].

In addition to individual functional genomics analyses, Schwartzentruber et al. [75] also developed a comprehensive gene prioritization score based on quantitative information derived from five predictors: gene distance to lead SNPs, eQTL colocalization, network score, bulk and single-cell gene expression, and the sum of fine-mapped probability for any coding SNPs within a gene. To identify weights for the predictive features, two models were employed and then the average of the predictions from the two models was used as the final gene prioritization “model score or probability” to determine the most likely functional genes for AD across GWAS loci. There were 12 most confidently prioritized genes with a model score or probability of  $> 0.80$ : *PILRA*, *APH1B*, *PLCG2*, *SPI1*, *SORL1*, *CD33*, *CASS4*, *BINI*, *CRI*, *ACE*, *ABCA7*, and *TSPAN14*; two additional genes, *PTK2B*, and *CD2AP*, had a model score of 0.79. With the exception of *CRI*, *CD2AP*, and *SPI1*, the other 11 genes overlap with genes identified from the eQTL colocalization and/or causal variants evidence. The latest and the largest

published GWAS/GWAS in 2021 [76] performed functional genomics analyses on 38 loci and, based on the position and eQTL information from brain and immune tissues, identified 989 genes. Further colocalization analysis identified a candidate causal gene in 9 of the 38 loci, including *TNIP1*, *MADD*, *APH1B*, *GRN*, *AC004687.2*, *ACE*, *NTN5*, *CD33*, and *CASS4*.

The success of the integration of GWAS variants with omics data depends on the size of the tissue-specific omics data and the relevancy of the available tissues and cells to AD. Evidence suggests that regulatory causal variants have modest and cell-type specific effects [120], and this requires a well-powered QTL data to detect modest effects [126]. Microglia is one of the most relevant cell type to AD; however, currently, microglia QTL data is available on a small number of brain tissues [75, 124], leading the investigators to also use peripheral myeloid cells (monocytes and macrophages) in AD functional genomics studies. A recent such study has integrated AD GWAS with multiple peripheral myeloid genomic datasets [125], which nominated multiple candidate genes: *BINI*, *SPI1*, *ZYX*, *EPHA1*, *MS4A6A*, *MS4A4A*, *PILRA*, *RABEP1*, *SCIMP*, *PTK2B*, *GPR141*, *SPPL2A*, and *CD2AP*. Noteworthy, the *EPHA1/EPHA1-AS1* association signal is an eQTL for the *EPHA1-AS1* noncoding RNA both in primary microglia and in IPSDMac [123], suggesting that the AD risk may be mediated by *EPHA1-AS1* expression levels.

Although the functional genomics analyses in each study summarized above was not done on the same set of genes and not all candidate genes in a given locus were examined, for now, the following have the strongest evidence to be considered as the causal genes for further cell-based and/or animal model studies: *APOE*, *ABCA7*, *BINI*, *TREM2*, *SORL1*, *SPI1*, *CRI*, *PTK2B*, *PILRA*, *CD2AP*, *APH1B*, *PLCG2*, *MS4A4A*, *MS4A6A*, *CD33*, *ADAM10*, and *ABI3*. Clearly, this is only a tentative list, which will certainly grow as functional genomics analyses are performed simultaneously on all known and GWS subthreshold genes on well-powered QTL data in AD-relevant tissues and cells.

## Concluding Remarks and Future Directions

Since 2009, the horizon of the AD genetics landscape has expanded enormously with the identification of at least 95 risk loci for LOAD, harboring both common and rare variants. With the projected availability of WGS data on large number of AD cases and controls in 2022, there is a high likelihood of discovering numerous addition rare variants in the existing and novel genes/loci, which certainly will lead to delineate the genetic architecture of AD. It is noteworthy, however, that the overall number of AD-associated loci remains low as compared with other neurodegenerative and

psychiatric disorders [127–129]. Like most other common diseases and traits, LOAD is also considered as a polygenic disease with the possible contribution of thousands of variants with small effect to its overall heritable risk [130]. However, recent data suggest that LOAD may be less polygenic than psychiatric diseases, and traits related to intelligence, cognitive ability, and educational attainment [131] or even it has an oligogenic architecture requiring only ~100 common SNPs [132]. Another recent study predicted the number of causal SNPs for LOAD to be about 11,000 [133]. These predictions are far less than the estimates of causal SNPs for other complex diseases and traits (Table 3).

Of the 58–70% reported heritability of LOAD [5, 6], the common SNP-based heritability from earlier known LOAD loci has been estimated to be up to ~35% [9–11, 131, 132], which needs to be re-evaluated based on the latest and larger identified loci. Some of the missing heritability may also be discovered using alternative approaches to the case–control design. Studies focusing on AD-related quantitative endophenotypes and biomarkers that manifest earlier than clinical AD provide a powerful alternative approach to identify not only additional AD-related genes, but it may also help to uncover underlying mechanisms for AD and disease progression that cannot be obtained from case–control studies. Some successful examples of this approach have been demonstrated in GWAS on AD biomarkers and endophenotypes, including amyloid and tau [134–137], neurofilament light (NFL)

[138, 139], neuropathologic features [140] resilience [141], psychosis [142], and AAO [132, 143–145], resulting in novel loci. Thus far, the focus of the vast majority of genomic studies in LOAD has been on European or European-derived White populations in overlapping subjects, despite the fact that the prevalence of AD is disproportionately higher in certain non-White groups. Special efforts need to be made to recruit and collect genetic material from a large number of non-White populations for meaningful and well-powered genetic association studies in order to identify potential novel ethnic-specific genetic factors, which is an essential step to complete the AD genetics landscape.

Even though the efforts to delineate the genetics landscape of AD continue, it is now time to divert attention and resources towards identifying the causal genes in all the identified loci so that the underlying disease causing biological mechanisms and pathways could be understood. A series of *in silico* functional genomics analyses have already begun to prioritize putative functional genes, but this effort has largely focused on study-specific-identified loci rather than on all the identified loci. Nevertheless, these studies have identified some promising candidate functional genes that will need to be tested further in cell-based systems and/or animal models to probe their roles in the pathogenesis of AD.

The ultimate goal of integrating genetic and functional findings is to discover novel pathways that may converge in the pathobiology of this heterogeneous disorder and eventually to identify drug targets for therapeutic treatment [146]. Due to the complexity of the underlying biological mechanisms and pathways causing the heterogeneity in the pathobiology of AD, a single drug target may not be therapeutic or it may work in only subset of patients. A combination drug treatment approach, as used for some other chronic diseases, has also been envisioned for AD [43]. In fact, anti-A $\beta$  and anti-tau therapies have been recommended to be used simultaneously [147] because both A $\beta$  and tau may act in parallel to exert their detrimental effects on AD [148, 149]. Despite the lack of clear success of anti-A $\beta$  monotherapy for symptomatic AD in clinical trials, recently, the US Food and Drug Administration (FDA) has granted authorization to the monoclonal antibody aducanumab as part of its “accelerated approval” pathway (treatments that may reasonably likely, but not certain to help patients), which has met with skepticism as well as hope [150, 151] because this is the first new drug for AD since 2003. It is highly likely that functional genomics-guided discoveries would lead to the identification of novel druggable targets and eventually therapeutic treatment of this devastating disease.

**Table 3** Estimates of causal SNPs for complex diseases/traits<sup>a</sup>

Disease/trait	Number of casual SNPs
Alzheimer's disease	99 <sup>b</sup>
Alzheimer's disease	11,200
Parkinson's disease	34,000 <sup>b</sup>
Major depression	173,000 <sup>b</sup>
Schizophrenia	185,000 <sup>b</sup>
Schizophrenia	582,000
Bipolar disorder	651,000
Crohn's disease	10,500
Ulcerative colitis	12,700
Coronary artery disease	20,700
Total cholesterol	12,700
LDL-cholesterol	64,300
HDL-cholesterol	27,900
Intelligence	140,000
Education	158,000
Body mass index	25,700

<sup>a</sup>Adapted from Holland et al. 2021 [133]

<sup>b</sup>From Zhang et al. 2020 [132]

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s13311-021-01152-0>.

**Acknowledgements** This study was supported by the National Institute on Aging (NIA) grants AG030653, AG041718, AG064877, and P30-AG066468. Technical help of Drs. Muaaz Aslam and Frank Fan is greatly appreciated in artwork and for providing useful feedback on genomic data.

**Required Author Forms** [Disclosure forms](#) provided by the authors are available with the online version of this article.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Hyman BT, Phelps CH, Beach TG, Bigio EH, Cairns NJ, Carrillo MC, et al. National Institute on Aging–Alzheimer’s Association guidelines for the neuropathologic assessment of Alzheimer’s disease. *Alzheimers Dement* 2012; 8:1-13.
- 2021 Alzheimer’s disease facts and figures. *Alzheimers Dement* 2021; 17:327-406.
- Alzheimer’s Disease International. World Alzheimer Report 2015 – updated in 2017.
- Chene G, Beiser A, Au R, Preis SR, Wolf PA, Dufouil C, Seshadri S. Gender and incidence of dementia in the Framingham Heart Study from mid-adult life. *Alzheimers Dement* 2015; 11:310-320.
- Gatz M, Reynolds CA, Fratiglioni L, Johansson B, Mortimer JA, Berg S, et al. Role of genes and environments for explaining Alzheimer disease. *Arch Gen Psychiatry* 2006; 63:168-174.
- Wingo TS, Lah JJ, Levey AI, Cutler DJ. Autosomal recessive causes likely in early-onset Alzheimer disease. *Arch Neurol* 2012; 69:59-64.
- Kamboh MI. Molecular genetics of late-onset Alzheimer’s disease. *Ann Human Genet* 2004; 68:381-404.
- Bertram L, McQueen MB, Mullin K, Blacker D, Tanzi RE. Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database. *Nat Genet* 2007; 39:17-23
- Ridge PG, Hoyt KB, Boehme K, Mukherjee S, Crane PK, Haines JL, et al. Assessment of the genetic variance of late-onset Alzheimer’s disease. *Neurobiol Aging* 2016; 200:e13-e20.
- Lee S, Harold D, Nyholt DR, ANZGene Consortium, International Endogene Consortium, Genetic and Environmental Risk for Alzheimer’s disease Consortium, et al. Estimation and partitioning of polygenic variation captured by common SNPs for Alzheimer’s disease, multiple sclerosis and endometriosis. *Hum Mol Genet* 2013; 22: 832–841.
- Brainstorm Consortium et al. Analysis of shared heritability in common disorders of the brain. *Science* 2018; 360: eaap8757.
- Bateman RJ, Xiong C, Benzinger TL, Fagan AM, Goate A, Fox NC, et al. Clinical and biomarker changes in dominantly inherited Alzheimer’s disease. *N Engl J Med* 2012; 367:795-804. Erratum in: *N Engl J Med* 2012; 367:780.
- Gordan BA, Blazey TM, Su Y, Hari-Raj A, Dincer A, Flores S, et al. Spatial patterns of neuroimaging biomarker change in individuals from families with autosomal dominant Alzheimer’s disease: a longitudinal study. *Lancet Neurol* 2018; 17:241-250.
- Jack CR Jr, Bennett DA, Blennow K, Carrillo MC, Dunn B, Haeberlein SB, et al. NIA-AA Research Framework: Toward a biological definition of Alzheimer’s disease. *Alzheimers Dement* 2018; 14:535-562.
- Zhu XC, Tan L, Wang HF, Jiang T, Cao L, Wang C, et al. Rate of early onset Alzheimer’s disease: a systematic review and meta-analysis. *Ann Transl Med* 2015; 3:38.
- Cacace R, Sleegers K, Van Broeckhoven C. Molecular genetics of early-onset Alzheimer’s disease revisited. *Alzheimers Dement* 2016;12:733-748.
- Jarmolowicz AI, Chen HY, Panegyres PK. The patterns of inheritance in early-onset dementia: Alzheimer’s disease and frontotemporal dementia. *Am J Alzheimers Dis Other Dement* 2015; 30:299–306.
- De Strooper, B. Vassar R, Golde T. The secretases: enzymes with therapeutic potential in Alzheimer disease. *Nat Rev Neurol* 2010; 6:99–107.
- Zhou R, Yang G, Guo X, Zhou Q, Lei J, Shi Y. Recognition of the amyloid precursor protein by human  $\gamma$ -secretase. *Science* 2019; 363:eaaw0930
- Xia W.  $\gamma$ -Secretase and its modulators: Twenty years and beyond. *Neurosci Lett* 2019; 701:162-169.
- Shah S Lee SF, Tabuchi K, Hao YH, Yu C, LaPlant Q et al. Nicastrin functions as a gamma-secretase substrate receptor. *Cell* 2005; 122:435–447.
- Wong E, Frost GR, Li YM.  $\gamma$ -secretase modulatory proteins: The guiding hand behind the running scissors *Front Aging Neurosci*. 2020; 12:614690.
- Hoogmartens J, Cacace R, Van Broeckhoven C. Insight into the genetic etiology of Alzheimer’s disease: A comprehensive review of the role of rare variants. *Alzheimers Dement (Amst)*. 2021; 13:e12155.
- Tew J, Goate AH. Genetics of  $\beta$ -amyloid precursor protein in Alzheimer’s disease. *Cold Spring Harb Perspect Med* 2017; 7:a024539.
- Selkoe DJ, Hardy J. The amyloid hypothesis of Alzheimer’s disease at 25 years. *EMBO Mol Med* 2016; 8:595-608.
- Sun L, Zhou R, Shi Y. Analysis of 138 PSEN1 pathogenic mutations in the presenilin-1 on the in vitro production of A $\beta$ 42 and A $\beta$ 40 peptides by  $\gamma$ -secretase. *Proc Natl Acad Sci USA* 2017; 114:E476-E485.
- Pottier C, Hannequin D, Coutant S, Rovelet-Lecrux A, Wallon D, Rousseau S, et al. High frequency of potentially pathogenic SORL1 mutations in autosomal dominant early onset Alzheimer disease. *Mol Psychiatry* 2012; 17:875-879.
- Nicolas G, Charbonnier C, Wallon D, Quenez O, Bellenguez C, Grenier-Boley B, et al. SORL1 rare variants: a major risk factor for familial early-onset Alzheimer’s disease. *Mol Psychiatry* 2016; 21:831-836.
- Kim M, Suh J, Romano D, Truong MH, Mullin K, Hooli B, et al. Potential late-onset Alzheimer disease associated mutations in the ADAM10 gene attenuate alpha-secretase activity. *Hum Mol Genet* 2009; 18:3987-3996.
- Shu J, Choi SH, Romano DM, Gannon MA, Lesinski AN, Kim DY, et al. ADAM10 missense mutations potentiate  $\beta$ -amyloid accumulation by impairing prodomain chaperone function. *Neuron* 2013; 80:385-40.
- Cruchaga, C. Haller G, Chakraverty S, Mayo K, Vallania FL, Mitra RD, et al. Rare variants in APP, PSEN1 and PSEN2 increase risk for AD in late-onset Alzheimer’s disease families. *PLoS ONE* 2012; 7:e31039.



32. Jonsson T, Atwal JK, Steinberg S, Snaedal J, Jonsson PV, Bjornsson S, et al. A mutation in APP protects against Alzheimer's disease and age-related cognitive decline. *Nature* 2012; 488:96-99.
33. Bamne MN, Demirci FY, Berman S, Snitz BE, Rosenthal SL, Wang X, 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* 2014; 35:1779.e1715–1779.e1716.
34. Wang LS, Naj AC, Graham RR, Crane PK, Kunkle BW, Cruchaga C, et al. Rarity of the Alzheimer disease-protective APP A673T variant in the United States. *JAMA Neurol* 2015; 72:209–216.
35. Xia Q, Yang X, Shi J, Liu Z, Peng Y, Wang W, et al. The Protective A673T mutation of amyloid precursor protein (APP) in Alzheimer's disease. *Mol Neurobiol* 2021; 58:4038-4050.
36. Kero M, Paetau A, Polvikoski T, Tanskanen M, Sulkava R, Jansson L, et al. Amyloid precursor protein (APP) A673T mutation in the elderly Finnish population. *Neurobiol Aging* 2013; 34:1518.e1511–1518.e1513.
37. Kamboh MI. Apolipoprotein E polymorphism and susceptibility to Alzheimer's diseases. *Hum Biol* 1995; 67:195-215.
38. Farrer LA, Cupples L, Haines JL, Kukull WA, Mayeux R, Pericak-Vance MA, 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* 1997; 278:1349–1356.
39. Seripa D, Matera MG, Daniele A, Bizzarro A, Rinaldi M, Gravina C, et al. The missing ApoE allele. *Ann Hm Genet* 2007; 71:496-500.
40. Mahley RW. Apolipoprotein E: from cardiovascular disease to neurodegenerative disorders. *J Mol Med (Berl)* 2016; 94:739-746.
41. Marais AD. Apolipoprotein E in lipoprotein metabolism, health and cardiovascular disease. *Pathology* 2019; 51:165-176.
42. Yamazaki Y, Zhao N, Caulfield TR, Liu C-C, Bu G. Apolipoprotein E and Alzheimer disease: pathobiology and targeting strategies. *Nat Rev Neurol* 2019; 15:501–518.
43. Long JM, Holtzman DM. Alzheimer Disease: An update on pathobiology and treatment strategies. *Cell* 2019; 179:312-339.
44. Lumdsden AL, Mulugeta A, Zhou A, Hyppönen E. Apolipoprotein E (APOE) genotype-associated disease risks: a phenome-wide, registry-based, case-control study utilising the UK Biobank. *EBioMedicine* 2020; 59:102954.
45. Chen Y, Strickland MR, Soranno A, Holtzman DM. Apolipoprotein E: Structural insights and links to Alzheimer disease pathogenesis. *Neuron* 2021; 109:205-221.
46. Genin E, Hannequin D, Wallon D, Sleegers K, Hiltunen M, Combarros O, et al. APOE and Alzheimer disease: a major gene with semi-dominant inheritance. *Mol Psychiatry* 2011; 16:903-907.
47. Liu M, Bian C, Zhang J, Wen F. Apolipoprotein E gene polymorphism and Alzheimer's disease in Chinese population: a meta-analysis. *Sci Rep.* 2014; 4:4383.
48. Rajabli F, Feliciano, BE, Celis K, Hamilton-Nelson, KL, Whitehead, PL, Adams LA, et al. Ancestral origin of ApoE ε4 Alzheimer disease risk in Puerto Rican and African American populations. *PLoS Genet* 2018; 14:21007791.
49. Kamboh MI, Aston CE, Perez-Tur J, Kokmen E, Ferrell RE, Hardy J, DeKosky ST. A novel mutation in the apolipoprotein E gene (*APOE\*4* Pittsburgh) is associated with the risk of late-onset Alzheimer's disease. *Neurosci Lett* 1999; 263:129-132.
50. Medway CW, Abdul-Hay S, Mims T, Ma L, Bisceglia G, Zou F, et al. ApoE variant p.V236E is associated with markedly reduced risk of Alzheimer's disease. *Mol Neurodegener* 2014; 9:11.
51. Blue EE, Cheng A, Chen S, Yu CE; Alzheimer's Disease Genetics Consortium. Association of uncommon, noncoding variants in the APOE region with risk of Alzheimer disease in adults of European ancestry *JAMA Netw Open* 2020; 3(10):e2017666.
52. Grupe A, Abraham R, Li Y, Rowland C, Hollingworth P, Morgan A, et al. Evidence for novel susceptibility genes for late-onset Alzheimer's disease from a genome-wide association study of putative functional variants. *Hum Mol Genet* 2007; 16: 865-873.
53. Coon KD, Myers AJ, Craig DW, Webster JA, Pearson JV, Lince DH, et al. A high-density whole-genome association study reveals that APOE is the major susceptibility gene for sporadic late-onset Alzheimer's disease. *J Clin Psych* 2007; 68: 613-618.
54. Reiman EM, Webster JA, Myers AJ, Hardy J, Dunckley T, Zismann VL, et al. GAB2 alleles modify Alzheimer's risk in APOE epsilon4 carriers. *Neuron* 2007; 54:713-720.
55. Li H, Wetten S, Li L, St Jean PL, Upmanyu R, Surh L, et al. Candidate single-nucleotide polymorphisms from a genomewide association study of Alzheimer disease. *Arch Neurol* 2008; 65: 45-53.
56. Bertram L, Lange C, Mullin K, Parkinson M, Hsiao M, Hogan MF, et al. Genome-wide association analysis reveals putative Alzheimer's disease susceptibility loci in addition to APOE. *Am J Hum Genet* 2008; 83: 623-632.
57. Beecham GW, Martin ER, Li YJ, Slifer MA, Gilbert JR, Haines JL, Pericak-Vance MA. Genome-wide association study implicates a chromosome 12 risk locus for late-onset Alzheimer disease. *Am J Hum Genet* 2009; 84: 35-43.
58. Abraham R, Moskvina V, Sims R, Hollingworth P, Morgan A, Georgieva L, et al. A genome-wide association study for late-onset Alzheimer's disease using DNA pooling. *BMC Med Genomics* 2008; 1:44.
59. Carrasquillo MM, Zou F, Pankratz VS, Wilcox SL, Ma L, Walker LP, et al. Genetic variation in PCDH11X is associated with susceptibility to late-onset Alzheimer's disease. *Nat Genet* 2009; 41:192-198.
60. Harold D, Abraham R, Hollingworth P, Sims R, Gerrish A, Hamshere ML, et al. Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nat Genet* 2009; 41:1088–1093.
61. Lambert JC, Heath S, Even G, Campion D, Sleegers K, Hiltunen M, et al. Genomewide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. *Nat Genet* 2009; 41:1094–1099.
62. Seshadri S, Fitzpatrick AL, Ikram MA, DeStefano AL, Gudnaso V, Boada M, et al. Genome-wide association analysis of genetic loci associated with Alzheimer disease. *JAMA* 2010; 303:1832–1840.
63. Hollingworth P, Harold D, Sims R, Gerrish A, Lambert JC, Carrasquillo MM, et al. Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. *Nat Genet* 2011; 43:429–435.
64. Naj AC, Jun G, Beecham GW, Wang LS, Vardarajan BN, Buros J, et al. Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. *Nat Genet* 2011; 43:436–441.
65. Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet* 2013; 45:1452–1458.
66. Kamboh MI, Demirci FY, Wang X, Minster RL, Carrasquillo MM, Pankratz VS, et al. Genome-wide association study of Alzheimer's disease. *Transl Psychiatry* 2012; 2:e117.
67. Jonsson T, Stefansson H, Steinberg S, Jonsdottir I, Jonsson PV, Snaedal J, et al. Variant of TREM2 associated with the risk of Alzheimer's disease. *N Engl J Med* 2013; 368:107–116.
68. Rosenthal SL, Bamne MN, Wang X, Berman S, Snitz BE, Klunk WE, et al. More evidence for association of a rare TREM2 mutation (R47H) with Alzheimer's disease risk. *Neurobiol Aging* 2015; 36:2443.e21e2443.e26.
69. Ruiz A, Heilmann S, Becker T, Hernandez I, Wagner H, Thelen M, et al. Follow-up of loci from the International Genomics of



- Alzheimer's Disease Project identifies TRIP4 as a novel susceptibility gene. *Transl Psychiatry* 2014; 4:e358.
70. Sims R, van der Lee SJ, Naj AC, Bellenguez C, Badarinarayan N, Jakobsdottir J, et al. Rare coding variants in *PLCG2*, *AB13* and *TREM2* implicate microglia-mediated innate immunity in Alzheimer's disease. *Nat Genet* 2017; 49:1373-1384.
  71. Liu JZ, Erlich Y, Pickrell JK. Case-control association mapping by proxy using family history of disease. *Nat Genet* 2017; 49:325-331.
  72. Marioni RE, Harris SE, Zhang Q, McRae AF, Hagenaars SP, Hill WD, et al. GWAS on family history of Alzheimer's disease. *Transl Psychiatry* 2018;8:99.
  73. Jansen IR, Savage JE, Watanabe K, Bryois J, Williams DM, Steinberg S, et al. Genome-wide meta analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat Genet* 2019; 51:404-413.
  74. Kunkle BW, Grenier-Boley B, Sims R, Bis JC, Damptte V, Naj AC, et al. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates  $A\beta$ , tau, immunity and lipid processing. *Nat Genet* 2019; 51:414-430.
  75. Schwartzenstruber J, Cooper S, Liu JZ, Barrio-Hernandez, I, Bello E, Kumasaka N, et al. Genome-wide meta-analysis, fine-mapping and integrative prioritization implicate new Alzheimer's disease risk genes. *Nat Genet* 2021; 53:392-402.
  76. Wightman DP, Jansen IE, Savage JE, Shadrin AA, Bahrami S, Holland D, et al. A genome-wide association study with 1,126,563 individuals identifies new risk loci for Alzheimer's disease. *Nat Genet* 2021; 53:1276-1282.
  77. Greaves CV, Rohrer JD. An update on genetic frontotemporal dementia. *J Neurol* 2019; 266:2075-2086.
  78. Li Z, Farias FHG, Dube U, Del-Aquila JL, Mihindukulasuriya KA, Fernandez MV, et al. The *TMEM106B* FTLN-protective variant, rs1990621, is also associated with increased neuronal proportion. *Acta Neuropathol* 2020; 139:45-61.
  79. Bellenguez C, Küçükali F, Jansen I, Andrade V, Moreno-Grau S, Amin N, et al. New insights on the genetic etiology of Alzheimer's and related dementia. medRxiv preprint 2020.10.01.20200659. <https://doi.org/10.1101/2020.10.01.20200659>. *Nat Genet* 2022; in press.
  80. Naj AC, Leonenko G, Jian X, Grenier-Boley B, Dalmasso C, Bellenguez C, et al. Genome-wide meta-analysis of late-onset Alzheimer's disease using rare variant imputation in 65,602 subjects identifies novel rare variant locus NCK2: The International Genomics of Alzheimer's Project (IGAP). medRxiv preprint 2021.03.14.21253553. <https://doi.org/10.1101/2021.03.14.21253553>.
  81. Jun G, Ibrahim-Verbaas CA, Vronskaya M, Lambert J-C, Chung J, Naj AC, et al. A novel Alzheimer disease locus located near the gene coding tau protein. *Mol Psychiatry* 2016; 21:108-117.
  82. Ma Y, Jun GR, Zhang X, Chung J, Naj AC, Chen Y, et al. Analysis of whole-exome sequencing data for Alzheimer disease stratified by APOE genotype. *JAMA Neurol* 2019; 76:1099-1108.
  83. Fan K-S, Feingold E, Rosenthal SL, Demirci FY, Ganguli M, Lopez OL, Kamboh MI. Whole-exome sequencing analysis of Alzheimer's disease in non-*APOE*\*4 carriers. *J Alzheimer Dis* 2020; 76:1553-1565.
  84. Reitz C, Jun G, Naj A, Rajbhandary R, Vardarajan BN, Wang LS, et al. Variants in the ATP-binding cassette transporter (*ABCA7*), apolipoprotein E  $\epsilon$ 4, and the risk of late-onset Alzheimer disease in African Americans. *JAMA* 2013; 309:1483-1492.
  85. Kunkle BW, Schmidt M, Klein HU, Naj AC, Hamilton-Nelson KL, Larson EB, et al. Novel Alzheimer disease risk loci and pathways in African American individuals using the African Genome Resources Panel: A meta-analysis. *JAMA Neurol* 2021; 78:102-113.
  86. Logue MW, Schu M, Vardarajan BN, Farrell J, Bennett DA, Buxbaum JD, et al. Two rare AKAP9 variants are associated with Alzheimer's disease in African Americans. *Alzheimers Dement* 2014; 10:609-618.
  87. Mez J, Chung J, Jun G, Kriegel J, Bourlas AP, Sherva R, et al. Two novel loci, *COBL* and *SLC10A2*, for Alzheimer's disease in African Americans. *Alzheimers Dement* 2017; 13:119-129.
  88. Reitz C, Mayeux R; Alzheimer's Disease Genetics Consortium. *TREM2* and neurodegenerative disease. *N Engl J Med*. 2013;369:1564-1565.
  89. Jin SC, Carrasquillo MM, Benitez BA, Skorupa T, Carrell D, Patel D, et al. *TREM2* is associated with increased risk for Alzheimer's disease in African Americans. *Mol Neurodegener* 2015; 10(1):19.
  90. Zhou X, Chen Y, Mok KY, Zhao Q, Chen K, Chen Y, et al. Identification of genetic risk factors in the Chinese population implicates a role of immune system in Alzheimer's disease pathogenesis. *Proc Natl Acad Sci USA* 2018; 115:1697-1706.
  91. Wang HZ, Bi R, Hu QX, Xiang Q, Zhang C, Zhang DF, et al. Validating GWAS-identified risk loci for Alzheimer's disease in Han Chinese populations. *Mol Neurobiol* 2016; 53:379-390.
  92. Jiang T, Tan L, Chen Q, Tan MS, Zhou JS, Zhu XC, et al. A rare coding variant in *TREM2* increases risk for Alzheimer's disease in Han Chinese. *Neurobiol Aging* 2016; 42:217.e1-3.
  93. Shigemizu D, Mitsumori R, Akiyama S, Miyashita A, Morizono T, Higaki S, et al. Ethnic and trans-ethnic genome-wide association studies identify new loci influencing Japanese Alzheimer's disease risk. *Transl Psychiatry* 2021; 11:51.
  94. Nalls M A, Pankratz N, Lill CM, Do CB, Hernandez DG, Saad M, et al. Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. *Nat Genet* 2014; 46:989-993.
  95. Redensek S, Trost M, Dolzan V. Genetic determinants of Parkinson's disease: can they help to stratify the patients based on the underlying molecular defect? *Front Aging Neurosci* 2017; 9:20
  96. Diabetes Genet. Replication Meta-Anal. (DIAGRAM) Consort., Asian Genet. Epidemiol. Netw. Type 2 Diabetes (AGEN-T2D) Consort., South Asian Type 2 Diabetes (SAT2D) Consort., Mex. Am. Type 2 Diabetes (MAT2D) Consort., Type 2 Diabetes Genet. Explor.Next-Gener. Seq. Multi-Ethnic Samples (T2D-GENES) Consort., et al. Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nat Genet* 2014; 46:234-244.
  97. Bien SA, Pankow JS, Haessler J, Lu YN, Pankratz N, Rohde RR, et al. Transethnic insight into the genetics of glycaemic traits: fine-mapping results from the Population Architecture Using Genomics and Epidemiology (PAGE) consortium. *Diabetologia* 2017; 60:2384-2398.
  98. Wojcik GL, Graff M, Nishimura KK, Tao R, Haessler J, Gignoux CR, et al. Genetic analyses of diverse populations improve discovery for complex traits. *Nature* 2019; 570:514-518.
  99. Bien SA, Wojcik GL, Hodonsky CJ, Gignoux CR, Cheng I, Matise TC et al. The future of genomic studies must be globally representative: Perspectives from PAGE. *Annu Rev Genom Hum Genet* 2109; 20:181-200.
  100. Jun GR, Chung J, Mez J, Barber R, Beecham GW, Bennett DA, et al. Transethnic genome-wide scan identifies novel Alzheimer's disease loci. *Alzheimer's Dementia* 2017; 13:727-738.
  101. Gurdasani D, Barroso I, Zeggini E, Sandhu MS. Genomics of disease risk in globally diverse populations. *Nat Rev Genet* 2019; 20:520-35.
  102. Kruglyak L. The road to genome-wide association studies. *Nat Rev Genet* 2008; 9: 314-318.
  103. Bis JC, Jian X, Kunkle BW, Chen Y, Hamilton-Nelson KL, Bush WS, et al. Whole exome sequencing study identifies novel rare and common Alzheimer's-Associated variants involved in immune response and transcriptional regulation. *Mol Psychiatry* 2020; 25:1859-1875. Published online 14 August 2018.

104. Brody JA, Morrison AC, Bis JC, O'Connell JR, Brown MR, Huffman JE, et al. Analysis commons, a team approach to discovery in a big-data environment for genetic epidemiology. *Nat Genet* 2017; 49:1560–1563.
105. The International HapMap Consortium. A second generation human haplotype map of over 3 million SNPs. *Nature* 2007; 449:851–861.
106. Hirschhorn JN. Genomewide association studies—illuminating biologic pathways. *N Engl J Med* 2009; 360:1699–1701.
107. Beecham GW, Bis JC, Martin ER, Choi SH, DeStefano AL, van Duijn CM, et al. The Alzheimer's Disease Sequencing Project: Study design and sample selection. *Neurol Genet* 2017; 3(5):e194.
108. Adekar SP, Klyubin I, Macy S, Rowan MJ, Solomon A, Dessain SK, O'Nuallain B. Inherent anti-amyloidogenic activity of human immunoglobulin gamma heavy chains. *J Bio Chem* 2010; 285:1066–1077.
109. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, et al. Systematic localization of common disease-associated variation in regulatory DNA. *Science* 2012; 337:1190–1195.
110. Farh KK, Marson A, Zhu J, Kleinewietfeld M, Housley WJ, Beik S, et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* 2015; 518:337–343.
111. Belkadi A, Bolze A, Itan Y, Cobat A, Vincent QB, Antipenko A, et al. Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. *Proc Natl Acad Sci USA* 2015; 112: 5473–5478.
112. Kirkpatrick D, Morgan SL, Mullin K, Hofmann O, Chapman B, Kirchner R, et al. Whole-genome sequencing reveals new Alzheimer's disease-associated rare variants in loci related to synaptic function and neuronal development. *Alzheimers Dement* 2021; 17:1509–1527.
113. Zhang B, Gaiteri C, Bodea L-G, Wang Z, McElwee J, Podtelezchnikov AA, et al. Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. *Cell* 2013; 153:707–720.
114. Mathys H, Davila-Velderrain J, Peng Z, Gao F, Mohammadi S, Young JZ, et al. Single-cell transcriptomic analysis of Alzheimer's disease. *Nature* 2019; 570:332–337.
115. Grubman A, Chew G, Ouyang JF, Sun G, Choo XY, McLean C, et al. A single cell brain atlas in human Alzheimer's disease. *Nat Neurosci* 2019; 22:2087–2097.
116. International Genomics of Alzheimer's Disease Consortium (IGAP). Convergent genetic and expression data implicate immunity in Alzheimer's disease. *Alzheimers Dement* 2015; 11:658–671.
117. Nott A, Holtman IR, Coufal NG, Schlachetzki JCM, Yu M, Hu R, et al. Brain cell type-specific enhancer–promoter interactome maps and disease-risk association. *Science* 2019; 366:1134–1139.
118. Gerring ZF, Gamazon ER, White A, Derks EM. Integrative network-based analysis reveals gene networks and novel drug repositioning candidates for Alzheimer disease. *Neurol Genet* 2021; 7:e622.
119. Porcu E, Rueger S, Lepik K, eQTLGen Consortium, BIOS Consortium, Santoni FA, et al. Mendelian randomization integrating GWAS and eQTL data reveals genetic determinants of complex and clinical traits. *Nat Commun* 2019; 10: 3300.
120. Gallagher MD, Chen-Plotkin AS. The post-GWAS era: From association to function. *Am J Hum Genet* 2018; 102:717–730.
121. Steinberg S, Stefansson H, Jonsson T, Johannsdottir H, Ingason A, Helgason H, et al. Loss-of-function variants in ABCA7 confer risk of Alzheimer's disease. *Nat Genet* 2015; 47:445–447.
122. Barbosa AC, Kim MS, Ertunc M, Adachi M, Nelson ED, McAnally J, et al. MEF2C, a transcription factor that facilitates learning and memory by negative regulation of synapse numbers and function. *Proc Natl Acad Sci USA* 2008; 105:9391–9396.
123. Young AMH, Kumasaka N, Calvert F, Hammond TR, Knights A, Panousis N, et al. A map of transcriptional heterogeneity and regulatory variation in human microglia. *Nat Genet* 2021; 53:861–868.
124. Corces MR, Shcherbina A, Kundu S, Gloude-mans MJ, Frésard L, Granja JM, et al. Single-cell epigenomic analyses implicate candidate causal variants at inherited risk loci for Alzheimer's and Parkinson's diseases. *Nat Genet* 2020; 52:1158–1168.
125. Novikova G, Kapoor H, TCW J, Abud EM, Efthymiou AG, Chen SX, et al. Integration of Alzheimer's disease genetics and myeloid genomics identifies disease risk regulatory elements and genes. *Nat Commun* 2021; 12:1610.
126. Yao D W, O'Connor L J, Price A L, Gusev A. Quantifying genetic effects on disease mediated by assayed gene expression levels. *Nat Genet* 2020; 52:626–633.
127. Nalls MA, Blauwendraat C, Vallerga CL, Heilbron K, Bandres-Ciga S, Chang D, et al. Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet Neurol* 2019; 18:1091–1102.
128. Lam M, Chen CY, Li Z, Martin AR, Bryois J, Ma X, et al. Comparative genetic architectures of schizophrenia in East Asian and European populations. *Nat Genet* 2019; 51:1670–1678.
129. Davies G, Lam M, Harris SE, Trampush JW, Luciano M, Hill WD, et al. Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function. *Nat Commun* 2018; 9:2098.
130. Escott-Price V, Sims R, Bannister C, Harold D, Vronskaya M, Majounie E, et al. Common polygenic variation enhances risk prediction for Alzheimer's disease. *Brain* 2015; 138:3673–3684.
131. Zhang Y, Qi G, Park J H, Chatterjee N. Estimation of complex effect size distributions using summary-level statistics from genome-wide association studies across 32 complex traits. *Nat Genet* 2018; 50:1318–1326.
132. Zhang Q, Sidorenko Q, Couvy-Duchesne B, Marioni RE, Wright MJ, Goate AM, et al. Risk prediction of late-onset Alzheimer's disease implies an oligogenic architecture. *Nat Commun* 2020; 11:4799.
133. Holland D, Frei O, Desikan R, Fan C-C, Shadrin AA, Smeland OB, et al. The genetic architecture of human complex phenotypes is modulated by linkage disequilibrium and heterozygosity. *Genetics* 2021; 217:iyaa046.
134. Deming Y, Li Z, Kapoor M, Harari O, Del-Aguila JL, Black K, et al. Genome-wide association study identifies four novel loci associated with Alzheimer's endophenotypes and disease modifiers. *Acta Neuropathol* 2017; 133:839–856.
135. Yan Q, Nho K, Del-Aguila JL, Wang X, Risacher SL, Fan KH, et al. Genome-wide association study of brain amyloid deposition as measured by Pittsburgh Compound-B (PiB)-PET imaging. *Mol Psychiatry* 2021; 26:309–321. doi:<https://doi.org/10.1038/s41380-018-0246-7>. Epub 2018 Oct 25.
136. Raghavan NS, Dumitrescu L, Mormino E, Mahoney ER, Lee AJ, Gao Y, et al. Association between common variants in RBFox1, an RNA-binding protein, and brain amyloidosis in early and pre-clinical Alzheimer disease. *JAMA Neurol* 2020; 77:1288–1298.
137. Damotte V, van der Lee SJ, Chouraki V, Grenier-Boley B, Simino J, Adams H, et al. Plasma amyloid  $\beta$  levels are driven by genetic variants near *APOE*, *BACE1*, *APP*, *PSEN2*: A genome-wide association study in over 12,000 non-demented participants. *Alzheimers Dement* 2021. <https://doi.org/10.1002/alz.12333>. Online ahead of print.
138. Niu LD, Xu W, Li JQ, Tan CC, Cao XP, Yu J, et al. Genome-wide association study of cerebrospinal fluid neurofilament light levels in non-demented elders. *Ann Transl Med* 2019; 7(22):657.
139. Wang ZT, Chen SD, Xu W, Chen KL, Wang HF, Tan CC, et al. Genome-wide association study identifies CD1A associated with rate of increase in plasma neurofilament light in non-demented elders. *Aging (Albany NY)* 2019; 11:4521–4535.
140. Beecham GW, Hamilton K, Naj AC, Martin ER, Huentelman M, Myers AJ, et al. Genome-wide association meta-analysis of neuropathologic features of Alzheimer's disease and related dementias. *PLoS Genet* 2014; 10(9):e1004606.

141. Dumitrescu L, Mahoney ER, Mukherjee S, Lee ML, Bush WS, Engelman CD, et al. Genetic variants and functional pathways associated with resilience to Alzheimer's disease. *Brain* 2020; 143:2561-2575.
142. DeMichele-Sweet MAA, Klei L, Creese B, Harwood JC, Weamer EA, McClain L, et al. Genome-wide association identifies the first risk loci for psychosis in Alzheimer disease. *Mol Psychiatry* 2021. <https://doi.org/10.1038/s41380-021-01152-8>. Online ahead of print.
143. Kamboh MI, Barmada MM, Demirci FY, Minster RL, Carrasquillo MM, Pankratz VS, et al. Genome-wide association analysis of age-at-onset in Alzheimer's disease. *Mol Psychiatry*. 2012; 17:1340-1346.
144. Naj AC, Jun G, Reitz C, Kunkle BW, Perry W, Park YS, et al. Effects of multiple genetic loci on age at onset in late-onset Alzheimer disease: a genome-wide association study. *JAMA Neurol* 2014; 71:1394-1404.
145. Huang KL, Marcora E, Pimenova AA, Di Narzo AF, Kapoor M, Jin SCA, et al. A common haplotype lowers PU.1 expression in myeloid cells and delays onset of Alzheimer's disease. *Nat Neurosci* 2017; 20:1052-1061.
146. Kamboh MI. A brief synopsis on the genetics of Alzheimer's disease. *Curr Genet Med Rep* 2018; 6:133-135.
147. Puzzo D, Argyrousi EK, Staniszewski A, Zhang H, Calcagno E, Zuccarello E. et al. Tau is not necessary for amyloid-beta-induced synaptic and memory impairments. *J Clin Invest* 2020; 130:4831-4844.
148. Pickett EK, Herrmann AG, McQueen J, Abt K, Dando O, Tulloch J, et al. Amyloid beta and tau cooperate to cause reversible behavioral and transcriptional deficits in a model of Alzheimer's disease. *Cell reports* 2019; 29:3592-604 e5.
149. Gulisano W, Maugeri D, Baltrons MA, Fa M, Amato A, Palmeri A, et al. Role of Amyloid-beta and Tau proteins in Alzheimer's disease: Confuting the amyloid cascade. *J Alzheimer's Dis* 2018;64(s1):S611-S631.
150. Mullard A. Landmark Alzheimer's drug approval confounds research community. *Nature* 2021; 594:309-310.
151. Walsh S, Merrick R, Milne R, Brayne C. Aducanumab for Alzheimer's disease? *BMJ* 2021; 374:n1682.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.