THE GENETIC ARCHITECTURE OF ALZHEIMER’S DISEASE: BEYOND APP, PSENs AND APOE

Rita J. Guerreiro¹,²,*, Deborah R. Gustafson³,⁴, and John Hardy⁵

¹Laboratory of Neurogenetics, National Institute of Aging, National Institutes of Health, Bethesda, Maryland, USA
²Center for Neuroscience and Cell Biology, University of Coimbra, Coimbra, Portugal
³Neuropsychiatric Epidemiology Unit, Sahlgrenska Academy, University of Gothenburg, Sweden
⁴Departments of Neurology and Medicine, SUNY-Downstate Medical Center, Brooklyn, New York, USA
⁵Reta Lila Weston Institute and Departments of Molecular Neuroscience and Neurodegenerative Disease, Institute of Neurology, London, United Kingdom

Abstract

Alzheimer’s disease (AD) is a complex disorder with a clear genetic component. Three genes have been identified as the cause of early onset familial AD (EOAD). The most common form of the disease is, however, a sporadic one presenting itself in later stages of life (LOAD). The genetic component of this late onset form of AD has been the target of a large number of studies, since only one genetic risk factor (APOE4) has been consistently associated with the disease. However, technological advances allow new approaches in the study of complex disorders. In this review, we discuss the new results produced by genome wide association studies, in light of the current knowledge of the complexity of AD genetics.

INTRODUCTION

Historically, Alzheimer’s disease (AD) is the most common cause of dementia. It is a progressive neurodegenerative disorder with an insidious onset, which typically appears in older individuals, but may affect people as early as the third decade of life [Rademakers, et al. 2003, Rogaeva 2002]. The genetics of Alzheimer’s disease is complex and heterogeneous. Most cases are “sporadic” with no apparent familial recurrence of the disease. However, a small percentage of AD cases (1–2% of all cases) have an early onset (EOAD), with symptoms appearing before 65 years of age. In these patients, the disease commonly aggregates within families and typically presents an autosomal dominant pattern of inheritance. Mutations in three genes are known to account for this early onset, familial
type of the disease: amyloid precursor protein gene (APP), presenilin 1 gene (PSEN1) and presenilin 2 gene (PSEN2) [Rogaeva 2002]. In fact, early-onset autosomal dominant disease with age of onset younger than sixty years, seems to be completely explained by pathogenic mutations in these three genes.

The most common, late onset and sporadic form of the disease remains mostly a genetic conundrum. The only well established genetic risk factor for late onset AD (LOAD) is the E4 allele of apolipoprotein E although many lifestyle risk factors have been reported including low education, mid-life high blood pressure and cholesterol levels, obesity and diabetes [Rogaeva 2002]. It has been estimated that the four established AD genes account for less than 30% of the genetic variance in EOAD and LOAD, suggesting that numerous additional AD genes may exist [Daw, et al. 2000]. The pursuit of these additional genes has been unproductive until recently. In the last two years however, high throughput technologies able to genotype up to one million single nucleotide polymorphisms (SNPs) have revealed some of the genetic players in different complex disorders. In this review, we will consolidate the current knowledge on the role of APP, PSENs and APOE in AD, we will systematically review the most promising new loci and variants reported to be associated with AD and we will discuss the recent advances in AD genome wide analysis in the context of the known epidemiology of the disorder.

**EOAD CAUSATIVE GENES (APP, PSEN1 and PSEN2)**

The amyloid precursor protein gene (APP, OMIM 104760, chromosome 21q21) encodes an ubiquitously expressed, integral type I membrane glycoprotein that exists as different alternatively spliced isoforms, with three predominant ones: APP751, APP770 and APP695, the latter being the main isoform found in brain [Yoshikai, et al. 1990]. The proteolytic processing of APP results in the production of different peptides (including Aβ), after a series of secretase cleavages, and occurs through two mutually exclusive pathways: the amyloidogenic pathway (fundamentally considered as the pathogenic pathway) and the non-amyloidogenic or constitutive pathway [Esch, et al. 1990, Haass, et al. 1992, Shoji, et al. 1992]. The identification of Aβ as a metabolic product of APP and the reports of AD families harboring APP causative mutations led to the general concept that Aβ is a key player in the development of AD, and that EOAD mutations are influencing the properties or ratios of the different Aβ isoforms in the brain [Hardy 1997]. Dominant mutations in APP are, however, a rare cause of AD with an estimated frequency of 16% of familial EOAD patients [Raux, et al. 2005]. More recently, two mutations in APP (A673V and E693Δ) have been reported to cause AD only in the homozygous state in families with apparently recessive modes of inheritance [Di Fede, et al. 2009] [Tomiyama, et al. 2008].

In addition to missense variants, copy number mutations have been identified in autosomal dominant early-onset families. Five French families [Cabrejo, et al. 2006, Rovelet-Lecrux, et al. 2006] were first reported to harbor small chromosomal duplications with different break points, but all including the APP locus. Subsequent screens in Finnish and Dutch AD cases revealed additional APP duplications in EOAD cases with prominent cerebral amyloid angiopathy (CAA) [Remes, et al. 2004, Rovelet-Lecrux, et al. 2007, Sleegers, et al. 2006]. The phenotypic spectrum of APP duplications is yet to be fully defined but clearly includes mixed phenotypes of AD and/or CAA. The estimated frequency of duplications also appears to be variable: in the selected Rovelet-Lecrux cohort it was 8% (about half the contribution of missense APP mutations to early onset, autosomal dominant AD) [Raux, et al. 2005]; in the Dutch cohort less than 2% [Sleegers, et al. 2006]; in EOAD familial and sporadic Swedish and Finnish cases there were no duplications in APP identified [Blom, et al. 2008]; and a frequency of 18% was estimated in early onset familial Japanese cases [Kasuga, et al. 2009].

*Neurobiol Aging*. Author manuscript; available in PMC 2013 March 1.
The presenilin 1 (PSEN1, OMIM 104311, chromosome 14q24.3) and presenilin 2 (PSEN2, OMIM 600759, chromosome 1q31-q42) genes have a very similar genetic structure and encode two proteins expressed in a multiplicity of tissues including the brain, with higher levels in the cerebellum and the hippocampus and a primarily neuronal expression [Levy-Lahad, et al. 1996],[Rogaev, et al. 1995]. These are highly homologous, sharing an overall amino acid sequence identity of 67%. Hydrophobicity plots predicted these to be integral membrane proteins [Rogaev, et al. 1995] most likely adopting a transmembrane structure containing nine segments with a hydrophilic intracellular loop region [Henricson, et al. 2005, Laudon, et al. 2005]. PSENs are important components of the multimeric gamma-secretase complex and are predominantly located in the endoplasmic reticulum and Golgi compartments, clearly suggesting their involvement in protein processing [Kovacs, et al. 1996],[De Strooper 2003]. The first disease causing mutations in PSEN1 and PSEN2 were identified in 1995 [Rogaev, et al. 1995, Sherrington, et al. 1995]. At the present time, 175 pathogenic mutations and 7 variants non-pathogenic or with unclear pathogenicity have been identified in PSEN1. PSEN2 harbors fewer mutations: 14 pathogenic mutations and 9 variants non-pathogenic or with unclear pathogenicity (http://www.molgen.ua.ac.be/ADMutations, accessed on August 2009). The PSENs mutation range encompasses mainly missense mutations, thus manifesting in a scattered fashion all over the proteins, with some clustering around transmembrane domains [Guerreiro, et al. 2008, Hardy and Crook 2001].

GENETIC RISK FOR AD (APOE)

The apolipoprotein E gene (APOE, OMIM 107741, chromosome 19q13.2) encodes a glycoprotein synthesized mainly in the liver, brain (primarily by neurons and astrocytes), and also by other cells such as macrophages and monocytes [Siest, et al. 1995]. APOE is involved in the mobilization and redistribution of cholesterol in the periphery and also during neuronal growth and repair [Mahley 1988]; in nerve regeneration, immunoregulation and activation of several lipolytic enzymes [Mahley and Rall 2000]. The three major APOE isoforms (ApoE2, ApoE3 and ApoE4) differ in two sites of the aminoacid sequence (residues 112 and 158) and are encoded by a single genetic locus. The frequencies of the ε2, ε3, and ε4 alleles were estimated at 0.11, 0.72, and 0.17, respectively, but vary widely among populations [Zannis, et al. 1981]. The ε4 allele, (the ancestral allele), is more frequent in populations such as Pygmies (0.407) and Khoi San (0.370). The ε2 allele frequency oscillates with no apparent trend and, for example, is absent in some Native Americans populations [Corbo and Scacchi 1999].

Many studies have demonstrated an association between the ε4 allele and familial and sporadic forms of LOAD. This allele represents an increased risk seen across different ethnic groups of 3 fold for heterozygous carriers and up to 15 fold for individuals who are ε4 homozygotes, when compared to ε3 homozygotes [Ashford 2004]. APOE is known to act in a dose dependent manner in AD: the effect of the ε4 allele in the risk for AD increases from 20 to 90% and the mean age of onset decreases from 84 to 68 years with the increase in the number of ε4 alleles [Corder, et al. 1993]. The ε2 allele has been shown to have an impact on longevity and may confer protection against AD [Corder, et al. 1994]. Distinct binding properties of the different APOE isoforms to the Aβ peptide [Strittmatter, et al. 1993] and tau protein [Strittmatter, et al. 1994] have been suggested to underlie the disparities associated with each genotype. In particular, the ApoE4 isoform binds to the Aβ peptide more rapidly than the ApoE3 isoform, forming novel monofibrils that precipitate into dense structures [Sanan, et al. 1994]. The fact that ApoE4 does not bind to tau protein in vitro, unlike ApoE2 and ApoE3, has suggested to some that this interaction between ApoE3 and tau serves as a protection against tau phosphorylation and consequent neurofibrillary tangle formation [Strittmatter, et al. 1994, Weisgraber 1994].
The ε4 allele appears to be a risk factor and not an invariant cause of AD, indicating that other environmental or genetic factors may need to be concurrently acting with this allele in order to cause AD [Hyman, et al. 1996]. For example, physical activity has been shown to be protective for dementia in non-APOEε4 carriers [Podewils, et al. 2005]. Additionally, while most of the risk at the APOE locus is likely to be encoded at the protein coding polymorphism, it is likely that other genetic variability at this locus, probably altering APOE expression, also contributes to the risk of developing AD [Bekris, et al. 2009, Chartier-Harlin, et al. 1994, Lambert, et al. 2002, Lambert, et al. 1997]. Genetic variability in APOE expression may contribute more to disease risk, rather than independent effects of the adjacent gene TOMM40.

During the past two decades there have been many studies searching for genetic risk factors: hundreds of positive and negative results have been produced, but none apart from APOE have produced clear and reproducible associations.

DIFFERENT TYPES OF STUDIES IN AD GENETICS

Methodologically, two main genetic strategies have governed the field: genetic linkage analysis and case-control association studies. For linkage analysis, researchers have used informative families where a clear heritability of the disease is present and no mutations have been found. In association studies, researchers compare the frequencies of a pre-determined allele between a group of AD cases and a group of healthy individuals. Several considerations are relevant for both strategies, of which we will mention the main ones. A full discussion of all methodological issues involved in these types of studies goes beyond the scope of this review. Here we will briefly describe the known genes and risk factors associated with AD, with a focus on genome wide association studies (GWAS) as a methodology for exploring genetic susceptibility and identifying risk factors and mechanisms for disease. We will also address the previously overlooked role of homozygosity and recessive cases in the major complexity of AD genetics.

LINKAGE STUDIES—Genetic linkage studies aim to identify chromosomal regions associated with disease by measuring the correlated segregation of particular markers with a determined phenotype within a family [Dawn Teare and Barrett 2005]. This type of study usually involves three sequential steps: 1) the identification of the disease causative locus; 2) sequencing the region found in the previous step in a cohort of cases and controls in order to define and characterize the mutation(s) found; 3) uncovering the molecular and biological functions of the genes found [Altshuler, et al. 2008]. Several factors are known to complicate this approach in AD: i) difficulties in getting large, complete and informative multigeneration families; ii) the potential inclusion of phenocopies (individuals with a sporadic indistinguishable form of disease); and iii) genetic heterogeneity, since observing that the pattern of disease in families is consistent with a major gene component does not necessarily imply that only one gene or factor is involved. Additionally, linkage mapping suffers from limitations, such as the low resolution of the results. Usually these studies do not identify one gene or one mutation associated with a disease, but rather a chromosomal region (many times, a very large region) is identified. In addition, the strongest linkage signals tend to come from recessive and highly penetrant, thus very rare, disorders [Dawn Teare and Barrett 2005].

Nonetheless, linkage mapping has been a very important methodology in the study of AD genetics. The four genes undoubtedly associated with AD were identified primarily by linkage analysis. In addition to these genes, several other genomic regions have been implicated using this methodology (Table 1). These loci contain many genes that have been considered candidates and consequently have been studied in order to identify the genetic
variation responsible for the development of AD. Until now, no specific genes implicated in AD have been identified in these regions. Most recently, a study by Butler and colleagues used a meta-analysis method to analyze the pooled linkage results from five independent genome scans. These included the results of 2206 affected individuals and 785 families of Caucasian and Caribbean Hispanic descent. This study was able to identify genome-wide suggestive evidence for linkage on chromosomes 1p13.3-q31.1, 7pter-p21.1 and 8p22-p21.1, together with other seven loci presenting nominally significant evidence for linkage [Butler, et al. 2009]. Interestingly, the most significant locus identified in this study (8p22-p21.1) includes the CLU gene, (the top hit from the largest GWAS performed in AD, as discussed below in this review) and previously reported loci by different studies (as 9p, 9q, 10q and 12p) were not identified by Butler and colleagues. Even so, linkage analysis has largely failed to identify risk factors in LOAD, probably due to the low odds ratios associated with the unidentified variants.

GENE ASSOCIATION STUDIES—The quest for genetic risk factors in clinical genetics has mainly focused on the study of candidate genes (usually focusing only on variants altering the coding sequence of a gene). These types of studies rely on a rather simple principle: to test if a determined allelic or genotypic variant occurs more or less often in a group of people with a particular disease when compared to a similar group of healthy individuals.

The success of this approach relies in an in-depth understanding of the disease and disease pathways, in such a way that the researcher will be able to select, not only the right gene, but also the right variant(s) to be studied. Additionally, the two studied groups need to be homogeneous, well characterized and large enough to allow a statistically powerful analysis [Hattersley and McCarthy 2005].

Most gene association studies in AD have studied a few variants in one or two genes. The large number of genes and even more variants, clearly reduce the chances of true positive findings. Nonetheless, several hundreds of positive associations have been reported. Most of these are certainly false positives resulting mainly from population substructure (i.e. existence of subpopulations in which there was random mating with reduced amount of gene flow, that may lead to spurious associations when the geographical origin/ancestry of cases and controls are not matched) [Tian, et al. 2008], poor statistical analysis and publication bias toward positive results [Choudhry, et al. 2006, Ioannidis, et al. 2001]. Positive associations have been broadly published, while negative results (unless convincingly refuting previous results) would not be reported. When reported, most of these studies pointed to small sample sizes, or specific genetic population backgrounds as main reasons for the negative results. Other issues, particularly in large epidemiologic studies, include case definition or the ability to accurately identify AD versus vascular or other forms of dementia; age of dementia onset estimation; and phenotypic variations in the disease whether cognitive, vascular, psychiatric or metabolic.

In order to address these very large numbers of conflicting reports, a database (the AlzGene database) was created which systematically collects, summarizes and meta-analyzes the results for all the genetic variants studied in association with AD [Bertram, et al. 2007]. As of November 16th 2009 the top ten results in this database included: APOE (E2/3/4), CLU, PICALM, TNK1, ACE, TFAM, CST3, IL1B, CR1 and hCG2039140 [Bertram and Tanzi 2008]. The fifth top hit, ACE, has repeatedly been reported as associated with AD [Webster J, et al. 2009], atherosclerosis [Sayed-Tabatabaei, et al. 2003] and hypertension [Staessen, et al. 1997]. The role of vascular risk factors in AD is further discussed below.
GENOME WIDE ASSOCIATION STUDIES—The development of platforms able to genotype millions of SNPs and of powerful analytical frameworks, able to distinguish true associations; together with the completion of the International Human HapMap project [Consortium 2005], have provided unprecedented tools to the study of the so called “Common Disease-Common Variant” (CD-CV) hypothesis. This theory proposes that common polymorphisms (usually defined as having a minor allele frequency of over 5%) may contribute to the overall susceptibility to common diseases [Cargill, et al. 1999, Chakravarti 1999, Lander 1996]. This type of analysis addresses one of the major pitfalls of gene association studies: the coverage of the study. Instead of studying one or two genetic variants, we are now able to test the majority of common variability in the genome for association with disease, by means of testing tagging SNPs, i.e., polymorphisms in linkage disequilibrium (LD) with each other. This means that if one knows the genotype in one locus, one can predict with a high accuracy (dependent on the strength of the LD and the allele frequencies) the genotype occurring at linked loci. However, these platforms require the analyses of large numbers of samples and some regions of the genome are still not well covered. Thus, the success of genome wide association studies (GWAS) depends on sample size, frequency of risk alleles and individual effect sizes. The smaller the attributable risk associated with any given common variant, the greater the number of samples needed to identify that variant. The risk allele frequency relates to the effect size: a risk variant that has a high odds ratio but is rare in the population is much more difficult to identify with GWAS than a common variant linked to moderate or mild risk within the population. Additionally, population stratification (when cases and non-cases are not genetically similar); population admixture (when several distinct, unrecognized sub-populations comprise the same group of individuals studied) and unreported relatedness in the population may also be problems if the study is not correctly designed [Simon-Sanchez and Singleton 2008].

The competing hypothesis to explain the genetic basis of complex diseases is the “Common Disease-Rare Variant” (CD-RV) theory, which suggests that multiple rare variants underlie susceptibility to such diseases [Fearnhead, et al. 2005]. Although the CD-CV versus CD-RV is a false dichotomy, it has major implications for the future of the research in this field, since different techniques are used to find common versus rare alleles. By definition, rare variants have low frequencies (MAF > 0.1% to 2–3%) and individually small contributions to the overall inherited susceptibility of a disease. In this way, rare variants will not be detectable by GWAS. Instead, the choice of candidate genes and appropriate case groups has been essential to uncover this type of variants[Bodmer and Bonilla 2008]. The new sequencing technologies will facilitate this approach by allowing the assessment of all the genes in the genome. Nonetheless, the extensive DNA resequencing of the whole genome in large numbers of individuals, and the assessment of the functional consequences of the variants found, poses difficult bioinformatic and data management challenges.

There is no doubt that GWAS have uncovered previously unknown polymorphic variants and genes with significant effects on AD risk. However, considering the studies so far carried out in AD (see next section for details), the associations still to be found will have ORs lower than 1.2. This has raised the question as to whether it is worthwhile to pursue even larger GWAS in order to identify variants with small effects disease risk. Certainly, in the future, sequencing and association strategies will be employed together to fully dissect the genetic architecture of the risk of AD and other complex disorders [Bodmer and Bonilla 2008].

Since APOE has been the only risk factor consistently associated with LOAD and some reports have estimated a heritability of AD between 60 and 80% [Gatz, et al. 2006, Pedersen, et al. 2001] it is believed that genetic variability plays a critical role in LOAD and that several risk factors are still to be uncovered. Daw and colleagues had estimated the
existence of four to seven additional genes contributing to this genetic variability [Daw, et al. 2000].

Several GWAS in AD have now been published. The first study of this type was performed by Grupe et al., in which the approach used was to first generate a short list of candidate SNPs by analyzing >17,000 SNPs in DNA pools from one set of cases and controls. In truth, this study only covers a small proportion of the genome, probably around 5%, and is not strictly “genome-wide”. In this study, markers meeting a previously defined significance criteria were then typed in DNA pools from a second set of samples and, again the SNPs satisfying the criteria were then individually genotyped in four sets of cases and controls and finally in a fifth sample. In total, nearly 4000 samples (~2000 AD cases, ~2000 controls) from the United States and the United Kingdom were studied. Three SNPs on chromosome 19 in LD to APOE represented the most significant associations, of which two (rs157581 in TOMM40 and rs405509 in APOE) achieved genome-wide significance. SNPs in other genes reached nominal significance but did not achieve genome wide significance [Grupe, et al. 2007].

The first studies covering the majority of the genome were published by Coon et al. [Coon, et al. 2007] and Reiman et al. [Reiman, et al. 2007] (~1000 cases and controls). In these studies on largely the same dataset, only a single SNP in LD with APOE (rs4420638) reached genome wide significance although SNPs in GAB2 (rs2373115) reached significance when the AD cases were stratified by APOE genotype [Reiman, et al. 2007]. An analysis of the top hits from AlzGene in the same dataset found that only SNPs is ACE reached nominal significance [Webster J, et al. 2009].

Liu et al. reported on a genome screen of 402 microsatellite markers in 103 LOAD patients and 170 first-degree relatives from a pedigree with 4645 members from The Netherlands. Multipoint analysis revealed four significant (1q21, 1q25, 10q22-24 and 3q22-24) and one suggestive (11q24-25) linkage peaks. Several of these regions coincide with previously reported loci (Table 1) with the strongest linkage association found for chromosome 1q21. Following these results, the authors tested for association between cognitive function and 4,173 SNPs in the linked regions in an independent sample consisting of 197 individuals from the same Genetic Research in Isolated Populations program. After adjusting for multiple testing, significant associations were identified in four (1q25, 3q22-24, 10q22-24 and 11q25) of the previous five regions (Table 2) [Liu, et al. 2007].

Abraham and colleagues performed a genome-wide association study in pooled DNA samples of ~2000 LOAD cases and >1000 controls. They identified a set of 109 SNPs with a significant association with AD and genotyped them individually. In addition to APOE, one SNP (rs727153), located approximately 13 kb apart from the start of LRAT transcription site, was suggested as associated with disease [Abraham, et al. 2008].

Li et al., 2008 studied a hypothesis-generating cohort from Canada and identified rs4420638 within APOCI to be strongly associated with AD, due to LD with APOE [Coon, et al. 2007] but nothing else reached genome wide significance [Li, et al. 2008].

Bertram et al., 2008 used samples from 410 AD families to identify five SNPs significantly or marginally associated with a multivariate phenotype combining age at onset of the disease and affection status. The marker presenting the most significant association was again rs4420638 (located 340 bp 3’ of APOCI) and almost certainly reflecting the effects of APOE ε4 allele. Once more, no other SNP reached genome wide significance [Bertram, et al. 2008].
Beecham and colleagues analyzed ~500 LOAD cases and ~500 cognitive controls followed by a further >200 cases and >200 controls used as a validation data set for SNPs that reached nominal, but not genome-wide significance. They too, were able to confirm association with the APOE locus and they suggest an association with the 12q13 locus, which they replicated in the validation data set. The associated SNP (rs11610206) is close to the hypothetical gene FAM113B. Additionally, there are a number of nearby candidate genes, such as VDR and AMIGO2. In order to validate associated SNPs with p values <0.0001 and nominally associated candidate genes, they imputed SNPs from a previously published GWAS [Reiman, et al. 2007]. Four additional highly associated signals were replicated with the use of the imputed data set: 1q42–within DISC1; 4q28–200 kb proximal to PCDH18; 6q14 – nearest gene BCKDHB; and 19q13 – within ZNF224 [Beecham, et al. 2009].

By analyzing ~1000 LOAD cases and ~1000 controls and evaluating the 25 SNPs with the most significant allelic association in four additional series, Carrasquillo and colleagues suggested an association with the X chromosome SNP rs5984894 in PCDH11X (Xq21.3) [Carrasquillo, et al. 2009] but again, this association did not reach genome wide significance.

Feulner et al. generated genome-wide data in a German cohort of ~500 AD patients and ~500 controls. The results obtained were analyzed only for the genes included in the top results list on the AlzGene database. Additionally to APOE, nominally significant associations were found for six of the ten studied genes (CH25H, PGBD1, LMNA, PCK1, MAPT and SORL1) [Feulner, et al. 2009] but none of these putative signals reached genome wide significance.

All the above studies, which each used <2000 Alzheimer cases in their analysis, were able to pick up the signal at the APOE locus, but despite tantalizing results, none were able to identify other loci at a level which passes the threshold for genome wide significance, and no two of them identified the same locus. The data from the Reiman et al. and by Li et al. studies, were made publicly available (http://www.tgen.org/research/neuro_gab2.cfm and http://www.GSK.com, respectively) to enable their additive use in other studies. Overall, these data pointed to the need for larger studies to identify risk loci with smaller effect sizes. Two such studies have now been reported.

One reported the analysis of ~4000 cases and ~8000 controls in stage 1 and ~2000 cases and ~2000 controls in stage 2. Two SNPs significantly associated with AD, outside the APOE locus, were identified: rs11136000 located in an intron of CLU on chromosome 8 and rs3851179 located close to the gene PICALM on chromosome 11. Although the risks associated with CLU and PICALM genes are relatively small (APOE odds ratio~4; CLU/PICALM~1.1) these associations reached genome wide significance [Harold, et al. 2009]. The other large AD GWAS reported on 2000 cases and ~5000 controls in the first stage, and ~3000 cases and ~3000 controls in the replication stage. An association between AD and ApoE, CLU and CR1 locus on chromosome 1q32 (OR=1.2) was found [Lambert, et al. 2009b].

These aforementioned three genes (CLU, CR1 and PICALM) will clearly be the subject of intense research. Here we will discuss the main features of each of these genes, as well as the potential pathobiological pathways in which they may be involved in the context of AD.

Clusterin or apolipoprotein J is, like APOE, a lipoprotein expressed in most mammalian tissues with higher levels present in brain, ovary, testis and liver [de Silva, et al. 1990]. CLU interacts with different molecules, including lipids, amyloid proteins, components of the complement membrane attack complex (MAC) and immunoglobulins [Jones and Jomary...
Accordingly, it has been proposed to be involved in a number of physiological processes such as ongoing synapse turnover [Danik, et al. 1993], apoptosis [Jenne and Tschopp 1992, Wong, et al. 1993], cytoprotection at fluid-tissue boundaries, membrane recycling during development and in response to injury and regulation of complement-mediated MAC [Jones and Jomary 2002, Oda, et al. 1994]. Clusterin has also been proposed to be a form of secreted heat-shock protein or chaperone molecule [Michel, et al. 1997, Wilson and Easterbrook-Smith 2000].

Several lines of evidence suggest that CLU has a central (either protective or pathogenic) role in the pathway leading to Alzheimer’s disease. First, CLU mRNA has been reported to be elevated in AD affected brain areas such as hippocampus, either when brains from AD patients were compared to one Huntington patient [Duguid, et al. 1989], or to controls [May, et al. 1990]. Likewise, Oda et al. reported a statistically significant difference in the Clusterin protein content of extracts from cortex and hippocampus when comparing 10 non-AD individuals and 25 AD patients [Oda, et al. 1994]. Second, Clusterin is one of the components of amyloid plaques [Kida, et al. 1995, McGeer, et al. 1992, McGeer, et al. 1994, Takamaru 1994]. Third, is able to bind soluble Aβ through a specific, reversible and high-affinity interaction in cerebrospinal fluid [Ghiso, et al. 1993, Golabek, et al. 1995] to form complexes able to cross the blood-brain barrier by a high affinity receptor mediated process involving transcytosis [Zlokovic 1996]. Fourth, reduced levels of APOE and increased levels of CLU have been correlated with the number of E4 alleles, suggesting a compensatory induction of CLU in the brain of AD individuals with the e4 allele of APOE presenting low brain levels of APOE [Bertrand, et al. 1995]. Moreover, CLU was shown to prevent aggregation and polymerization of synthetic Aβ and to enhance the oxidative stress caused by A in vitro [Matsubara, et al. 1996, Oda, et al. 1995], and to facilitate Aβ uptake in cell culture experiments [Hammad, et al. 1997]. Clusterin appears to regulate the toxicity and conversion of Aβ into soluble forms [Boggs, et al. 1996, DeMattos, et al. 2002, Matsubara, et al. 1996, Oda, et al. 1995]. Together with APOE, suppresses Aβ deposition [DeMattos, et al. 2004] and may modify Aβ clearance at the blood brain barrier [Bell, et al. 2007].

PICALM, encodes the phosphorylcholine-binding clathrin assembly protein, also known as CALM: clathrin assembly lymphoid-myeloid leukemia gene. It is ubiquitously expressed with particularly high levels in neurons. This gene has been associated with leukemia, thus its relation to AD may appear not as direct as the one observed for CLU (Figure 1). Nonetheless, its involvement in clathrin-mediated endocytosis (essential to the intracellular trafficking of proteins and lipids) [Kim and Kim 2001] and in the fusion of synaptic vesicles to the presynaptic membrane by directing the trafficking of VAMP2 [Harel, et al. 2008] have lead Harold et al. to propose two interesting hypotheses for the role of Picalm in AD. In this way, genetic variability in PICALM may result on synapse perturbations, possibly through synaptic vesicle cycling, or on alterations of APP processing through endocytic pathways, culminating in changes in Aβ levels [Harold, et al. 2009].

CR1, the complement component (3b/4b) receptor 1 (Knops blood group) is a member of the receptors of complement activation family. The gene encodes a monomeric single-pass type I membrane glycoprotein found on erythrocytes, leukocytes, glomerular podocytes, and splenic follicular dendritic cells that mediates cellular binding to particles and immune complexes that have activated complement [Ahearn and Fearon 1989].

Three complement pathways are known: the classical, alternative and lectin-mediated cascades, which have different activation triggers, but all terminate with the production of the membrane attack complex. High enough concentrations of MAC result in cell lysis. This may lead to tissue damaging when the complement activation tight regulation, (that occurs
through the action of several different endogenous complement inhibitor proteins), is deficient [Sjoberg, et al. 2009]. Typically, the complement classical pathway is the one associated with AD [Akiyama, et al. 2000], mainly due to three facts: 1) C1q, the first protein in this pathway, efficiently binds to aggregated Aβ, activating the pathway and further enhancing Aβ aggregation and fibril formation [Rogers, et al. 1992, Webster, et al. 1995]; 2) early complement activation proteins (C1q, C4 and C3) and the MAC have been found to co-localize with senile plaques, NFTs and dystrophic neurites in AD brains [McGeer, et al. 1989, Veerhuis, et al. 1996, Webster, et al. 1997]; and 3) increased mRNA levels of complement proteins are present in AD brains when compared to controls [Walker and McGeer 1992]. More recently, activation of the alternative pathway in AD brains has also been demonstrated: Aβ activates the alternative pathway in vitro [Bradt, et al. 1998]; factor B mRNA is present in AD frontal cortex, and factor D cleaved split products of factor B (Bb and Ba) are significantly increased in AD brains [Strohmeyer, et al. 2000]. After the discovery that the complement system can be activated in the brain by several senile plaques and neurofibrillary tangle related components, in the absence of antibodies, and that neurons are a source of complement proteins in the brain, the involvement of the complement system in AD has been widely accepted [McGeer and McGeer 2001]. However, whether this involvement has a protective or deleterious effect has been extensively debated. In fact, it has been proposed that binding of C1q to misfolded proteins in early AD, together with C4BP that decreases MAC activation, are favorable and enable clearance of the misfolded material. But, when the system is overwhelmed by amyloid, this protein binds extensively to C1q leading to the full activation of the complement, ultimately leading to detrimental inflammation and neurodegeneration [Sjoberg, et al. 2009].

In summary, as in other aging degenerative diseases, the complement system has an important role in AD, and one may expect that the activation of the classical or the alternative pathways (or both) by Aβ will lead to neurodegeneration in individuals with a genetic predisposition [Zipfel 2009]. This will possibly result from an unbalance between the expression of regulator proteins, and one or more cascade proteins (Figure 2). This model is able to explain, at least partially, the presence of neuropathological changes in the brains of non-demented individuals [Hof, et al. 1996] since the genetic variability in the complement genes may be responsible for different complement reactions to the presence of NFT and senile plaques. Interestingly, as mentioned above, one function attributed to clusterin is in the regulation of complement-mediated membrane attack complex. Together with vitronectin, clusterin binds to the nascent amphiphilic C5b-9 complex, rendering it water soluble and lytically inactive, raising the possibility that the genetic risk conferred by clusterin for the development of AD, may arise from its regulation role in the complement system.

**VASCULAR RISK FACTORS AND AD**

These genetic findings point to the role of tissue and vascular damage in AD pathogenesis and thus to vascular risk factors in its aetiology. It is noteworthy that many of the genes now implicated potentially have a direct role at the blood brain interface: this includes APOE, ACE and the complement cascade components including CR1 and CLU. As such, these findings are consistent with the epidemiologic literature, which has consistently reported an association between vascular risk factors and AD. The evidence base for the prevention of AD and related brain pathologies is strongest for control of vascular risk factors [Gustafson D and I 2009]. Overweight and obesity is a cornerstone of vascular risk, which predisposes to hypertension, hypercholesterolemia, diabetes, and cardiovascular disease. Indeed, the low risk ratios observed in genetic studies attempting to identify new susceptibility genes for AD may be due to lack of information on presence and/or severity of this vascular involvement, as well as to the differential expression and clustering of vascular and metabolic traits. AD
brain pathologies exist against a background continuum of vascular pathologies, which may
modulate risk for clinically manifest disease. Vascular factors most directly related to newly
identified susceptibility genes are hypertension, hypercholesterolemia, and overweight and
obesity.

Hypertension is a risk factor for stroke, ischemic white matter lesions, silent infarcts, general
atherosclerosis, myocardial infarction and cardiovascular morbidity and mortality. This risk
increases with increasing blood pressure also at apparently healthy blood pressure ranges
[Kannel 2000]. Several longitudinal studies have suggested an association between AD and
al. 1996, Stewart, et al. 2009]. ACE, which is currently (accessed on November 16th) the
fifth top hit on Alzgene database [Bertram, et al. 2007], plays a classical role in blood
pressure regulation as part of the renin-angiotensin system [Goossens, et al. 2003]. This
system may play a role in dementia pathogenesis because of its effects on vascular and
metabolic homeostasis, as well as amyloid metabolism. The gene encoding for ACE is an
AD susceptibility gene whereby effect modification is observed by vascular phenotype,
particularly with population stratification by vascular factors such as APOE ε4 allele,
systolic blood pressure, body mass index, and obesity indices [Gustafson, et al. 2008,
Katzov, et al. 2004]. Thus, the renin angiotensin system may also provide a link between
obesity, hypertension, and vascular syndromes, such as type 2 diabetes, and health of the
express the renin-angiotensin system [Strazzullo, et al. 2003].

Cholesterol is important in AD, not only because of its relationship with cardiovascular
disease, but due to its role in amyloid metabolism [Sparks, et al. 1994]. APOE and CLU are
two proteins involved in lipid transport in the peripheral and central nervous systems
[Nuutinen, et al. 2009]. APOE regulates cholesterol homeostasis in astrocytes and microglia,
and is related to blood cholesterol levels [Hoglund and Blennow 2007]. In addition,
mutations in the APP gene upregulate Aβ40 and Aβ42 production; [Pangalos, et al. 2005]
and Aβ processing is sensitive to cholesterol levels and lipid trafficking. Brain cholesterol
levels increase during AD progression [Hoglund and Blennow 2007]. The epidemiologic
evidence associating high blood cholesterol levels with AD and other forms of dementia is
mixed. High cholesterol levels in mid-life may increase risk for subsequent dementia and
low cholesterol levels have been predictive of subsequent dementia [Mielke, et al. 2005,
Reitz, et al. 2004] or no association has been observed [Li, et al. 2005, Yoshitake, et al.
1995]. Nevertheless, even within the mid-life cholesterol literature, results are conflicting, as
some studies have not found high cholesterol to predict later dementia [Kalmijn, et al. 2000,
Stewart, et al. 2007, Tan, et al. 2003]. While there remain a number of questions regarding
the amyloid hypothesis in relationship to AD, the potential link to cholesterol metabolism
and vascular damage is noteworthy [Hardy 2009].

Finally, while overweight and obesity appear to increase risk for dementia independently of
other vascular factors, there is limited evidence related to adipose-specific mechanisms of
action in AD, particularly in relationship to these newly identified susceptibility genes. Mid-
life total or central obesity measured decades before dementia onset has been linked to
in the eighth decade of life have also been observed [Gustafson, et al. 2003, Hayden, et al.
2006]. During the prodromal phase of dementia, higher rates of body weight or BMI decline
Stewart, et al. 2005]. Thus, one cannot deny the role of excess adiposity as enhanced
substrate for CR1-related inflammatory events, nor its potential role in hypertension and
hyperlipidemia. However, also of importance are the implications of declining metabolic parameters, such as BMI [Barrett-Connor, et al. 1996, Buchman, et al. 2005, Stewart, et al. 2005], blood pressure [Stewart, et al. 2009] and cholesterol [Stewart, et al. 2007] in AD, for which these newly identified susceptibility genes may enhance our precision in identifying subgroups of AD for whom interventions are more advantageous.

POSSIBLE ROLE OF HOMOZYGOSITY AND RECESSIVE CASES

AD, as many other diseases occurring sporadically, recurs within families more often than expected by chance alone. However, in the majority of the cases, the pattern of familial recurrence is not compatible with simple Mendelian transmission and this model is typically presumed to reflect a multifactorial determination with contributions from multiple genes and/or environmental factors [Altshuler, et al. 2008]. However, the observed familial recurrence could also be attributed to genetic loci with large phenotypic effects and reduced penetrance (possibly recessive loci). In this case, one would not necessarily expect to see recurrence of the disease in multiple generations, nor a high recurrence rate among siblings, and the disease would be sporadic in the population. Although without a definite confirmation of pathogenicity, two rare potentially disease-associated mutations (Q170H and R181G) in ADAM10 (an alpha-secretase capable of anti-amyloidogenic proteolysis of the amyloid precursor protein) were recently reported to be associated with LOAD [Kim, et al. 2009]. Other mutations in PSEN1 (A79V) [Kauwe, et al. 2007] and PSEN2 (N141I) [Levy-Lahad, et al. 1995] have also been reported to be present in families with non-carriers affected individuals.

Recessive contributions can be inferred when populations with high degrees of consanguinity present higher prevalence of the disease than the general population [Mani, et al. 2002]. The Wadi Ara population is one example of this premise: an unusually high prevalence of AD (20% of those over 65 years and 60% of those over 85 years) in a population where the 4 allele of APOE is relatively uncommon [Bowirrat, et al. 2000, Bowirrat, et al. 2001]. In this regard, the study of families with a recessive mode of inheritance may not only identify the cause of disease in the respective family, but also be of utility in the identification of risk factors contributing to the sporadic form of disease.

Populations that have been largely isolated and subjected to extensive inbreeding during considerable periods in their recent history represent a powerful resource for the study of new genetic variants for common diseases. These populations provide several advantages for genetic research, such as longer stretches of linkage between neighboring markers, high levels of genetic and environmental homogeneity and a simpler genetic architecture for complex traits. Although these long homozygous tracts of uninterrupted sequences may represent deletion polymorphisms, loss of heterozygosity or segmental uniparental disomy, recent data suggests that these, most likely, represent autozygosity (homozygosity by virtue of parental descent from a common ancestor) [Devilee, et al. 2001, Li, et al. 2006, Raghavan, et al. 2005, Woods, et al. 2004]. In this way, an obvious application of whole genome platforms in relation to autozygosity is in the genetic analysis of consanguineous families. This can be considered as an analogous approach to linkage analysis, in which researchers aim to define shared regions of autozygosity and/or overlapping structural variants in order to determine the role of autozygosity in a particular disease. Autozygosity mapping has long been recognized as a rapid and cost effective way to identify loci underlying recessive disease [Lander and Botstein 1987] and several genes underlying different disorders have been identified using this methodology [Camargos, et al. 2008, Paisan-Ruiz, et al. 2009]. Specifically in AD, we have generated the first catalog of autozygosity in EOAD by studying a consanguineous Israeli family. Although we were unable to pinpoint a specific gene due to the small number of samples we had available for...
analysis, we were able to generate a catalog that may be used in future studies of other families [Clarimon, et al. 2008].

Furthermore, even in outbred populations, more individuals have a high frequency of these autozygous tracts than previously expected [Gibson, et al. 2006, Li, et al. 2006]. Our group and others have recently reported the unexpected high degree of apparent parental consanguinity in control individuals from North America (~10% of studied individuals harboring homozygous tracts larger than 5Mb) [Simon-Sanchez, et al. 2007]. Similar numbers (~6%) were presented by Li and collaborators, when studying an outbred population of unrelated Han Chinese [Li, et al. 2006] and by Gibson et al. who reported that >1000 tracts exceeding 1Mb in length were observed in the ~200 unrelated HapMap individuals studied [Gibson, et al. 2006]. These observations prompted us to study autozygosity in LOAD in an outbred population. By comparing measures of extended homozygosity (greater than 1 Mb in length) in a population of >800 LOAD cases and >550 controls we were able to identify one homozygous region on chromosome 8 (8p12, not including the CLU locus), significantly associated with LOAD. Additionally, the comparison of the total numbers of homozygous runs and the total length of these runs between cases and controls, revealed a suggestive difference in these measures (p-values 0.052–0.062), most likely symptomatic of a recessive component in the etiology of LOAD [Nalls, et al. 2009]. The role of recessive mutations in AD has been considerably overlooked: in addition to the recent works describing recessive APP mutations (discussed above in this review) [Di Fede, et al. 2009, Tomiyama, et al. 2008], only two other studies of isolated populations with a high incidence of the disease have been reported where the primary analyses were performed using dominant or additive modes of inheritance [Farrer, et al. 2003, Liu, et al. 2007]. The new technologies now available will allow us to overcome this gap in the near future.

THE NEXT STEPS

Two major and clearly non-exclusive pathways have been recently discussed as the future guidelines in the genome wide analysis of complex disorders [Goldstein 2009, Hirschhorn 2009, Kraft and Hunter 2009]. The first is the extension of the assembly of studies containing larger (tens or even hundreds of thousands) and representative samples in order to identify variants with lower frequencies that may have been missed until now and that could explain the so elusive fractions of “missing heritability” in AD. The increase in the number of studied samples will inevitably result in the discovery of new variants and probably new genes associated with AD, but the real net value of these variants is highly disputable. Therefore, the actual dilemma now is to know how far one should take these studies in order to keep a positive balance between the resources applied and the gathered genetic returns [Goldstein 2009, Hardy and Singleton 2009]. Clearly, still to discover rare variants and variants with small effect sizes will be difficult to replicate, due to reduced power and restriction to specific populations, respectively. Nonetheless, in order to identify these rare variants one may speculate the need to use new chips with a better coverage of rare variants and the resequencing of previously identified regions. One may also predict that, being AD a complex disorder, more emphasis will be put in the study of endophenotypes, as happened in the study by Liu et al. that used cognitive function as an endophenotype of AD and identified the RGLS2, RALGPS2 and C1orf49 genes as potential causative genes located in one of the associated genomic regions [Liu, et al. 2007]. This approach will obviously require precise and clearly defined clinical assessments.

The second line of investigation will rely in the sequencing of whole exomes and whole genomes, expectantly unveiling several new rarer risk variants. These higher risk variants may be related to the commoner lower risk variants found with GWAS. In fact, a classical
example comes from the study of Parkinson's disease genetics: rare high-risk variants in the α-synuclein gene are the cause of monogenic PD [Polymeropoulos, et al. 1997], while a common haplotype of this same gene has been established as a moderate risk cause of sporadic disease [Simon-Sanchez, et al. 2009].

Both lines of research will be followed by large resequencing efforts, in order to identify the real risk variants. Many GWAS have identified regions of the genome associated with AD, but in many cases the real risk/causal changes are not known. Additionally, in the cases where SNPs have been associated with a disease, these SNPs identified may in fact be in LD with the real variants causing the association. Further functional data, (for instance, the effects of a determined variant in gene expression), although not always possible to obtain, will be essential not only to validate the previously identified variants, but also because many variants map to non coding protein sequences, gene deserts, or genomic regions without any functional elements. The integration of these functional studies with information on multiple variants from the same gene, in different populations and the effects of epigenetics and epistasis will be vital. The interpretation of these results, however, will be complex and will most likely require the combination of disciplines as integrative genomics and systems biology. Although difficult, this approach will ultimately allow a more profound understanding of the molecular pathways underlying AD and AD risk, as well as the subsequent identification of effective biomarkers and drugs.

Acknowledgments

This work was supported in part by the Intramural Research Program of the National Institute on Aging, National Institutes of Health, Department of Health and Human Services; project number Z01 AG000950-06 and Fundacao para a Ciencia e Tecnologia, Portugal grant SFRH/BD/27442/2006. DRG was supported by the Swedish MFR and JH by an MRC Returning Scientist Award.

REFERENCES


*Neurobiol Aging*. Author manuscript; available in PMC 2013 March 1.


Neurobiol Aging. Author manuscript; available in PMC 2013 March 1.
genome association analysis shows that ACE is a risk factor for Alzheimer’s disease and fails to replicate most candidates from Meta-analysis. Int J Mol Epidemiol Genet. 2009 (in press).


Figure 1. Predicted interactions for Clu (A), Picalm (B) and CR1 (C)
The software STRING 8.0 (available at http://string.embl.de/newstring_cgi/show_input_page.pl?UserId=75H_IKjjgP5Xi&sessionId=9vx3bNEK6tmB) was used to establish a network of predicted interactions for Clu (A), Picalm (B) and CR1 (C) proteins. Accessed on August 2009. STRING is a database of known and predicted protein interactions. The interactions include direct (physical) and indirect (functional) associations; they are derived from four sources: genomic context, high-throughput experiments, (conserved) coexpression and previous knowledge. STRING quantitatively integrates interaction data from these sources for a large number of organisms, and transfers information between these organisms where applicable. The database currently covers 2,483,276 proteins from 630 organisms. From the three represented proteins, CLU is the one with a more direct relation with AD: APOE is directly connected to CLU. Interestingly, CLU also interacts with proteins present in CR1 network (like C3). No other genes consistently associated with AD are present in PICALM or CR1 networks, indicating that these proteins may be involved in new pathobiological pathways.
Figure 2. Genetic variability and the role of the complement system in AD

Aβ and probably NFTs are able to activate the classical and alternative complement pathways. These are regulated by several membrane (m) and soluble (s) proteins at different stages. The genetic variability in CR1 is now known to be associated with the risk of developing AD. Increased numbers of samples are needed to know if the same is true for any other component of these pathways. Drawn from the work of Tenner [Tenner 2001] and McGeer and McGeer [McGeer and McGeer 2002]
## Table 1

Ten most interesting AD linkage regions.

<table>
<thead>
<tr>
<th>Chr</th>
<th>Region</th>
<th>Number of studies</th>
<th>Studies</th>
<th>LOD scores</th>
<th>Relevant genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1p13.3-q23.3</td>
<td>2</td>
<td>[Liu, et al. 2007]; [Butler, et al. 2009]</td>
<td>5.2</td>
<td>GSTM4; GSTM1; GSTM3; CSF1; NFG; HMGC5; PRKAB2; APH1A; CTSS; THEM5; FAM63A;</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CHRN B2; LMNA; PMVK; FDPS; APOA1BP; GBA; NTRK1; CRP; NCSTN</td>
</tr>
<tr>
<td>1</td>
<td>1q23.3-q31.1</td>
<td>3</td>
<td>[Liu, et al. 2007]; [Butler, et al. 2009]; [Blacker, et al. 2003]</td>
<td>2.1 – 4.0</td>
<td>F11R; USF1; FCER1G; RGS4; APOA2; RXRG; POU2F1; PRDX6; SOAT1; PTGS2</td>
</tr>
<tr>
<td>8</td>
<td>8p22-p21.1</td>
<td>2</td>
<td>[Butler, et al. 2009]; [Lee, et al. 2008]</td>
<td>&gt;2.0</td>
<td>NAT1; NAT2; LPL; ADRA1A; CHRNA2; CLU</td>
</tr>
<tr>
<td>10</td>
<td>10q21-q22</td>
<td>7</td>
<td>[Curtis, et al. 2001]; [Olson, et al. 2002]; [Myers, et al. 2002]; [Blacker, et al. 2003]; [Holmans, et al. 2005]; [Liu, et al. 2007]; [Hamshere, et al. 2007]</td>
<td>1.8 – 4.15</td>
<td>ZWINT; UBE2D1; TFAM; BIC1; ANK3; CDC2; EGR2; CTNNA3; LRRCTM; DNASJ12; SIRT1; SRG1; SUPV3L1; TSPAN15; VPS26A; HK1; TACR2; NEUROG3; NAR1A; SGPL1; PSAP; CHST3; PPICB; SEC24C; NDST2; CAMK2G; PLAU; VCL; AP3M1; MYST4; KCNMA1</td>
</tr>
<tr>
<td>12</td>
<td>12p11-p13</td>
<td>4</td>
<td>[Pericak-Vance, et al. 1997]; [Curtis, et al. 2001]; [Myers, et al. 2002]; [Holmans, et al. 2005]</td>
<td>1.4 – 3.9</td>
<td>TNSR5FIA; CNAP1; GAPDH; GNB3; C1R; APOBEC1; MMP3; A2M; PZP; A2MP; OLR1; LRP6; GRIN3B; GYS2; ARCC9; PKP2P1</td>
</tr>
<tr>
<td>21</td>
<td>21q21-q22</td>
<td>4</td>
<td>[Olson, et al. 2002]; [Myers, et al. 2002]</td>
<td>1.6 – 4.5</td>
<td>PRSS7; NCAM2; APP; C21orf63; C21orf55; RUNX1; C21orf55; DYRK1A;</td>
</tr>
</tbody>
</table>
These regions were selected either from the top 3 regions resulting from the recent meta-analysis performed by Butler and colleagues, or from the overlapping between four or more whole-genome studies assessing linkage with AD risk, based on the AlzGene database of concordant linkage/association regions observed in full genome screens (http://www.alzforum.org/res/com/gen/alzgene/linkage.asp). The relevant genes for each of the represented regions were obtained from crosschecking with the AlzGene database (www.alzgene.org, [Bertram, et al. 2007]).

Note that: results by [Kehoe, et al. 1999] and [Lee, et al. 2004] were not considered as they were significantly extended in the same groups’ follow-up genome screen (Myers, et al, 2002 and Lee et al., 2006, respectively). The results by Hamshere, et al., 2007 consist of an amalgamated of three data sets previously studied by Kehoe, et al., 1999; Myers, et al., 2002; Blacker, et al., 2003 and Holmans, et al., 2005. Other studies have been published that performed re-analyses of existing data either by incorporating other variables (such as evidence for AD with and without psychosis [e.g. [Bacanu, et al. 2002], [Avramopoulos, et al. 2005], [Hollingworth, et al. 2007]), or parent of origin effects e.g. [Bassett, et al. 2002]. Results by [Ashley-Koch, et al. 2005] and [Hahs, et al. 2006] on Amish families were not considered due to the inclusion of MCI cases.
Table 2

Main features and results of genome wide association studies performed in AD.

<table>
<thead>
<tr>
<th>Study</th>
<th>Platform</th>
<th>#SNPs</th>
<th>Population</th>
<th>Genotype data publicly available</th>
<th># Subjects</th>
<th>Replication stage(s)</th>
<th>Significant results</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Grupe, 2007</strong> [Grupe, et al. 2007]</td>
<td>Celera (cSNPs)</td>
<td>17343</td>
<td>UK and USA</td>
<td>No</td>
<td>380</td>
<td>4 replication tiers (UK2, 309 cases+349 controls; UK3, 503 cases+643 controls; WU, 376 cases+344 controls; SD, 240 cases+330 controls)</td>
<td>rs157581 (TOMM40, chr 19q13.32) rs405509 (APOE, chr 19q13.32) rs3745833 (GALP, chr 19q13.42) rs1554948 (TNK1, chr 17p13.1) rs1132899 (APOC2, chr 19q13.32) rs11622883 (14q32.13) rs8192708 (PCK1, chr 20q13.31) rs505058 (LMNA, chr 1q22) rs3800324 (PGBD1, chr 6p22.1) rs0907175 (LOC651924, chr 6q24.1) rs859849 (7p15.2) rs41310885 (hCV2227461, FAM63A, chr 1q21.2) rs3074877 (MYH13, chr 17p13.1) rs41271951 (bCV15746640, CTSS, chr 1q21.2) rs440103 (UBD, chr 6p22.1) rs9608099 (BCR, chr 22q11.23) rs2882676 (ACAN, chr 15q26.1) rs13022344 (TRAK2, chr 2q33.1) rs13016976 (EBF3, chr 10q26.3)</td>
</tr>
<tr>
<td><strong>Coon, 2007</strong> [Coon, et al. 2007]</td>
<td>Affymetrix (500K)</td>
<td>502627</td>
<td>USA, Netherlands; overlaps with Reiman, 2007</td>
<td>No</td>
<td>664</td>
<td>446 Neurpathological replication cohort (197 cases + 114 controls); Clinical replication cohort (218 cases + 146 controls) Total number of samples = 415 cases + 260 controls</td>
<td>rs901104; rs1385600; rs1007837; rs2510038; rs945261; rs7101429; rs10793294; rs4291702; rs7115850; rs2373115 (GAB2, chr 11q14.1)</td>
</tr>
<tr>
<td><strong>Reiman, 2007</strong> [Reiman, et al. 2007]</td>
<td>Affymetrix (500K)</td>
<td>312316</td>
<td>USA, Netherlands; overlaps with Coon, 2007</td>
<td>Yes</td>
<td>446</td>
<td>290 Neurpathological replication cohort (197 cases + 114 controls); Clinical replication cohort (218 cases + 146 controls) Total number of samples = 415 cases + 260 controls</td>
<td>rs901104; rs1385600; rs1007837; rs2510038; rs945261; rs7101429; rs10793294; rs4291702; rs7115850; rs2373115 (GAB2, chr 11q14.1)</td>
</tr>
<tr>
<td><strong>Liu, 2007</strong> [Liu et al. 2007]</td>
<td>Affymetrix (500K)</td>
<td>262000</td>
<td>Netherlands</td>
<td>No</td>
<td>4173</td>
<td>4173 SNPs in 197 unrelated subjects from the GRIP population</td>
<td>Chr 1q25 (RGSL2, RALGPS2, Clorf069); Chr 3q22-24 (NMMAT3, CLSTN2); Chr 10q22-24 (HTR7, MPHOSPH1, CYP2C); Chr 11q25 (OPCML, HNT)</td>
</tr>
<tr>
<td>Study</td>
<td>Platform</td>
<td>#SNPs</td>
<td>Population</td>
<td>Genotype data publicly available</td>
<td># Subjects</td>
<td>Replication stage(s)</td>
<td>Significant results</td>
</tr>
<tr>
<td>-------------------</td>
<td>-----------------------------------</td>
<td>---------</td>
<td>-----------------------------------</td>
<td>----------------------------------</td>
<td>------------</td>
<td>----------------------</td>
<td>-------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Li, 2008</td>
<td>Affymetrix (500K)</td>
<td>469438</td>
<td>Canada and UK</td>
<td>Yes</td>
<td>753</td>
<td>736</td>
<td>rs10868366; rs7019241 (GOLM1, chr 9q21.33); rs4420638 (APOC1, chr 19q13.32) also significant in Coon, 2007; rs10519262 (chr 15q21.2); rs9886784 (chr 9p24.3)</td>
</tr>
<tr>
<td>Abraham, 2008</td>
<td>Illumina HumanHap300 + Illumina Sentrix HumanHap240 S</td>
<td>561494</td>
<td>UK</td>
<td>No</td>
<td>1082 LOAD pooled</td>
<td>1239 controls pooled</td>
<td>114 SNPs were individually genotyped in the cases and controls used in the pools and 1400 controls were added ApoE; rs727153 (~13 kb from the start of transcription of LRAT, chr 4q32.1), also showed increased significance (10 fold) when additional controls were combined</td>
</tr>
<tr>
<td>Bertram, 2008</td>
<td>Affymetrix 500K</td>
<td>484522</td>
<td>Self-reported European descent from NIMH Genetics Initiative Study sample</td>
<td>No</td>
<td>1376 AD from 410 families</td>
<td>-</td>
<td>4 SNPs in 3 independent AD family samples: NIA- 1040 samples from 329 families; NCRAD- 1108 samples from 331 families; CAG- 483 samples from 215 sibships (Total: 2689 samples 1816 affected and 845 unaffected) Rs4420638 (located 340 bp 3' of APC1 and probably reflects the effects of apoE E4 allele); rs11159647 (in predicted gene NT_026437.1360, chr 1q31.2), also significant in replication samples and in TGEN’s data; rs179943 (in ATXN1, chr 6p22.3), trend for significance in replication samples; rs3826656 (in predicted gene NT_011109.848, chr 19q13.33), also significant in replication samples; rs2049161 (in cdNA BC040718, chr 18p11.31)</td>
</tr>
<tr>
<td>Beecham, 2009</td>
<td>Illumina HumanHap550</td>
<td>532000</td>
<td>USA</td>
<td>No</td>
<td>492 LOAD</td>
<td>496</td>
<td>1 SNP in 238 cases + 220 controls ApoE; rs11610206 (downstream of FAM113B, chr 12q13), also significant in replication sample 1q2; 4q28; 6q14; 19q13</td>
</tr>
<tr>
<td>Carrasquillo, 2009</td>
<td>Illumina HumanHap300</td>
<td>313504</td>
<td>USA</td>
<td>No</td>
<td>844 LOAD</td>
<td>1255</td>
<td>25 SNPs in 1547 cases + 1209 controls ApoE; rs5984894, rs2573905 (in PCDH11X, chr Xq21.3), also significant in replication sample</td>
</tr>
<tr>
<td>Study</td>
<td>Platform</td>
<td>#SNPs</td>
<td>Population</td>
<td>Genotype data publicly available</td>
<td># Subjects</td>
<td>Replication stage(s)</td>
<td>Significant results</td>
</tr>
<tr>
<td>---------------</td>
<td>----------------------------------</td>
<td>-------</td>
<td>------------</td>
<td>----------------------------------</td>
<td>------------</td>
<td>----------------------</td>
<td>-------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Harold, 2009</td>
<td>Illumina 610 quadchip (Illumina HumanHap550 and HumanHap300 in some samples)</td>
<td>529205</td>
<td>Europe, USA</td>
<td>Yes</td>
<td>3941</td>
<td>7848</td>
<td>2 significant SNPs in 2023 cases + 2340 controls ApoE; rs11136000 (in CLU, chr 8p21-p12), also significant in replication sample rs385179 (5' of PICALM, chr 11q14, also significant in replication sample)</td>
</tr>
<tr>
<td>Lambert, 2009</td>
<td>Illumina Human 610-Quad BeadChip</td>
<td>537029</td>
<td>France, Europe</td>
<td>No</td>
<td>2032</td>
<td>5328</td>
<td>Significant SNPs in 3978 probable AD cases + 3297 controls from Belgium, Finland, Italy and Spain rs11136000, rs279590, rs9331888 (in CLU, chr 8p21-p12), also significant in replication sample rs6656401 (in CR1, chr 1q32), also significant in replication sample</td>
</tr>
</tbody>
</table>