

Chapter 4

Genetic Risk Factors for Alzheimer's Disease

Onofre Combarros

Abstract The rare and familial early-onset alzheimer's disease (EOAD) is linked to fully penetrant mutational genes, whereas the commonest nonfamilial and late-onset alzheimer's disease (LOAD) is the result of multiple susceptibility genes, each one contributing a small amount to the total risk. Smaller-scale genetic association studies identified the e4 allele of the apolipoprotein E (APOE) gene as the best established LOAD risk factor, increasing risk by approximately 4 or 15 times for one or two e4 alleles, respectively. Hundreds of other candidate genes have been tested for association with LOAD, and meta-analyses of conflicting results were collected in the AlzGene database. Instead of studying only a few genetic variants in small sample sizes, larger-scale genetic association studies (genome-wide association study or GWAS) make it possible to evaluate essentially all genes and intergenic regions, in large international consortia with sufficient number of cases and controls. The four largest LOAD GWAS consortia joined forces forming a mega-consortium known as the International Genomics of Alzheimer's Project (IGAP) and conducted a mega-meta-analysis of 25,500 LOAD cases and 48,500 unaffected controls. In addition to APOE e4 allele, IGAP identified 19 susceptibility loci, but the effects of all these genes on LOAD risk are exceedingly small, increasing or decreasing the risk by approximately 1.30 times, at most. It is critical to investigate the functional basis for these LOAD-associated GWAS loci and their influence on gene expression (mRNA profiling). Examining the influences of these loci on endophenotypes (cerebrospinal fluid biomarkers and neuroimaging and cognitive measures) can help to predict age at onset and rate of preclinical and clinical progression of LOAD. The proportion of heritability of LOAD unexplained by GWAS findings could be due to rare variants that may be identified by whole exome and whole genome sequencing. In addition, a part of the still elusive genetic variability in LOAD could be due to gene-gene interaction effects.

O. Combarros, MD, PhD
Neurology Service and School of Medicine, University Hospital "Marqués de Valdecilla",
University of Cantabria, Santander 39008, Spain
e-mail: combarro@unican.es

Keywords Alzheimer's disease • Risk genes • Genome-wide association study
Meta-analysis • Whole exome sequencing • Epistasis

Early-Onset alzheimer's Disease and Late-Onset alzheimer's Disease: Causative and Susceptibility Genes

Multiple genetic defects involving either predictive (mutational) or susceptibility (risk) genes have been linked to the development of Alzheimer's disease (AD) [1–3]. Rare (1–2 % of all AD cases) and fully penetrant disease-causing mutations in three different genes (*APP*, amyloid-beta protein precursor; *PSEN1*, presenilin 1; and *PSEN2*, presenilin 2) lead to early-onset (patients younger than 65 years) Mendelian (familial) forms of AD (EOAD). Of note, mutations in these three genes explain disease in only about 13 % of patients with EOAD [4]. The vast majority of AD cases, the so-called sporadic AD with no apparent familial recurrence, are defined by onset age later than 65 years or late-onset AD (LOAD), and this LOAD form does not carry Mendelian-causing mutations but is believed to be the result of multiple risk genes which do not reliably cause the disease but increase an individual's susceptibility or predisposition to developing AD. Susceptibility genes are associated with the risk of LOAD, but each one contributes only a small amount to the risk. Twin studies predicted the heritability of LOAD to be as high as 80 % [5]. Susceptibility genes are identified by genetic association studies in which allele frequency for single-nucleotide polymorphisms (SNPs) at or near a gene is compared between AD cases and controls. Susceptibility genes are revealed when case and control frequencies differ significantly. There appears to be no overlap between the genes driving Mendelian versus non-Mendelian form of the disease; that is, common SNPs in *APP*, *PSEN1*, and *PSEN2* do not seem to contribute to risk for LOAD [6].

Smaller-Scale Genetic Association Studies: The Candidate Gene Approach

Apolipoprotein E

The candidate gene approach was successful for identifying the $\epsilon 4$ allele of the apolipoprotein E (*APOE*) gene on chromosome 19q13, as the only gene variant considered to be an “established” LOAD risk factor [7]. Unlike the EOAD mutations that are fully penetrant, *APOE* $\epsilon 4$ allele is a genetic risk factor that is neither necessary nor sufficient for the development of LOAD [1, 3, 4, 8]. Whereas only 24–30 % of the general Caucasian population carries at least one *APOE* $\epsilon 4$ allele, 40–65 % of LOAD patients have at least one copy of this allele. That is, many

LOAD patients have no *APOE* $\epsilon 4$ allele, and individuals carrying this allele may never develop LOAD, suggesting that there are additional factors modulating the influence of the *APOE* $\epsilon 4$ allele in causing the development of LOAD. People with one *APOE* $\epsilon 4$ allele have a roughly four times increased risk of AD, and those with two *APOE* $\epsilon 4$ alleles have a roughly 15 times increased risk, compared with those with the most common genotype *APOE* $\epsilon 3/3$ [4]. Although there is evidence of a risk effect of *APOE* $\epsilon 4$ in non-Europeans, the estimated effect sizes are smaller with less consistent results in African American and Hispanic subjects, which may suggest different underlying genetic or environmental factors for these ethnic groups. The effect of *APOE* $\epsilon 4$ appears to be age dependent: the lifetime risk for LOAD in individuals with the *APOE* $\epsilon 4/4$ genotype is high, estimated as 33 % for men or 32 % for women by the age of 75 years; by age 85 years, the risk climbs to 52 % for men and 68 % for women [9]. This very high-risk estimates for *APOE* $\epsilon 4$ carriers seem similar to those associated with autosomal dominant Mendelian genes. Therefore, *APOE* has been proposed as a moderately penetrant gene with semi-dominant inheritance: not all *APOE* $\epsilon 4$ carriers develop disease (hence, the $\epsilon 4$ allele in this gene is not fully penetrant), and heterozygous *APOE* $\epsilon 4$ carriers have intermediate risk compared with homozygous carriers [9]. Considering the delayed penetrance of LOAD, lack of preventive therapies, and the potential for psychological harm, genetic testing for *APOE* is not recommended. However, when LOAD prevention becomes possible, thus, this recommendation will need to be reconsidered, and genetic testing might be indicated for either high-risk groups (e.g., family members of LOAD cases) or for population screening.

Inconsistent Replication: Meta-analyses in the AlzGene Database

Since the original report of *APOE* as a genetic risk factor for LOAD in 1993, hundreds of genes have been tested for association with LOAD and reported in the literature. Most candidate gene association studies in LOAD have studied a few variants in only one or two genes, and despite positive initial results, inconsistent replication of original association findings has been the rule rather than the exception (except for *APOE*) and even for candidate genes with convincing functional data and thorough genetic assessment [1]. Multiple testing, population stratification, genotyping errors, and initial small sample size are potential reasons for false-positive findings in the original study. In addition, underpowered studies that are too small to detect a modest effect size can lead to false-negative follow-up studies. Candidate gene association studies have revealed modest estimated effect sizes with odds ratios (ORs) of less than 2.0 for risk alleles or greater than 0.5 for protective alleles. It is estimated that thousands to tens of thousands of subjects are required to have sufficient power to detect such effect sizes, a prerequisite that has typically not been fulfilled for many association studies in LOAD [10]. To address these very large numbers of conflicting results, a database (the AlzGene database) was created

which systematically collected, summarized, and meta-analyzed the results for all the genetic variants studied in association with LOAD [11]. As of 18 April 2011, the 10 genes with the strongest signals for association in the database AlzGene included *APOE* and nine other candidate genes (*BINI*, *CLU*, *ABCA7*, *CRI*, *PICALM*, *MS4A6A*, *MS4A4E*, *CD33*, and *CD2AP*), all of which came from genome-wide association studies.

Larger-Scale Genetic Association Studies: The Genome-Wide Approach

The HapMap Project and the 1000 Genomes Project

Instead of studying one or two genetic variants, recent advances now make it possible to evaluate essentially all genes and all regions between genes in a single experiment, a method called genome-wide association study (GWAS). The GWAS method represented an important advance compared to “candidate gene” studies in which sample sizes were generally smaller and the variants assayed were limited to a selected few associations that were difficult to replicate. The International HapMap Project [12, 13] launched in October 2002 led to the generation of a database of the common variants (defined as minor allele frequency of greater than 5 %) and the underlying linkage disequilibrium (LD) structure, or correlation between neighboring SNPs, providing the foundation for the GWAS. GWAS uses tagging SNPs, for example, polymorphisms in LD with each other, and 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 [14]. Understanding LD not only allows the construction of haplotypes but also provides the ability to impute the genotypes of nearby unobserved (not genotyped) SNPs using directly observed genotypes. Imputing facilitates merging data from different genotyping platforms with incomplete overlap [10]. Until 2010, GWAS studies had almost exclusively employed the HapMap data set as the reference panel for imputation of their genetic data, which contained up to two to three million SNPs. Using genome-wide sequencing with high-throughput platforms, the 1000 Genomes Project Consortium [15] described the location, allele frequency, and local haplotype structure of approximately 15 million SNPs, 1 million short insertions and deletions, and 20,000 structural variants. Over 95 % of the currently accessible variants found in any individual are present in this data set. From 2010 onward, the 1000 Genomes Project has increased power of GWAS to detect genetic influences due to less common variants. Rigorous quality control and statistical methods coupled with sufficient sample size can lead to high reproducibility of GWAS. Disadvantages of GWAS are that signals can be in intergenic regions making assessment of the functional relevance difficult, genetic methods often cannot identify which single-nucleotide variant is pathogenic, and most signals are from small effect loci [14].

International Consortia: Meta-analyses of GWAS

A consensus has emerged that a P value less than 5×10^{-8} corresponds to genome-wide significance in a non-African population-based GWAS. This is a conservative Bonferroni correction based on roughly one million “effectively independent” common SNPs throughout the genome. This involves the risk of rejecting biologically valid hypotheses on purely statistical grounds, that is, false negatives. Therefore, statistical power is the main threat to GWAS, necessitating the formation of large international consortia that can provide sufficient number of cases and controls. The four largest LOAD GWAS consortia are the European Alzheimer's Disease Initiative (EADI) based in France, the US-based Alzheimer's Disease Genetics Consortium (ADGC), the Genetic and Environmental Risk in Alzheimer's Disease (GERARD) group from the UK, and the neurology subgroup of the multinational Cohorts for Heart and Aging in Genomic Epidemiology (CHARGE) consortium. The first two GWAS were published in 2009 by the GERARD [16] and EADI [17] consortia. In approximately 6,000 LOAD and 10,000 control subjects, in addition to APOE-related SNPs that revealed genome-wide significance ($P=4.9 \times 10^{-37}$ to 1.8×10^{-157}), the GERARD consortium found that rs11136000 in clusterin (*CLU*, $P=8.5 \times 10^{-10}$, OR=0.86) and rs3851179 in the phosphatidylinositol-binding clathrin assembly protein (*PICALM*, $P=1.3 \times 10^{-9}$, OR=0.86) were significantly associated with LOAD. Analyzing 6,000 LOAD and more than 8,000 control subjects from EADI consortium, rs11136000 in *CLU* and rs6656401 in complement component receptor 1 (*CRI*, $P=3.7 \times 10^{-9}$, OR=1.21) were significantly associated with LOAD. In 2010, in more than 35,000 persons, the CHARGE consortium reported strong evidence that rs744373 near bridging integrator 1 gene (*BINI*, $P=1.59 \times 10^{-11}$, OR=1.13) was significantly associated with LOAD [18]. In 2011, two simultaneously published manuscripts reported meta-analyses of the findings of the ADGC, CHARGE, GERARD, and EADI consortia and described strong evidence for five new LOAD risk loci. In nearly 20,000 cases and 40,000 controls, Hollingworth et al. [19] described association with LOAD of rs3764650 in *ABCA7* ($P=5.0 \times 10^{-21}$, OR=1.23), rs610932 in *MS4A6A* ($P=1.2 \times 10^{-16}$, OR=0.91), rs9349407 in *CD2AP* ($P=8.6 \times 10^{-9}$, OR=1.11), rs11767557 in *EPHA1* ($P=6.0 \times 10^{-10}$, OR=0.90), and rs3865444 in *CD33* ($P=1.6 \times 10^{-9}$, OR=0.91). In approximately 19,000 cases and 29,000 controls, Naj et al. [20] confirmed that common variants at *MS4A* gene cluster, *CD2AP*, *CD33*, and *EPHA1* were associated with LOAD.

International Genomics of Alzheimer's Project: Mega-meta-analysis of GWAS

The four LOAD GWAS consortia have joined forces, forming a mega-consortium known as the International Genomics of Alzheimer's Project (IGAP). The project drew on data from a total of 74,000 people of European ancestry (25,500 LOAD and 48,500 unaffected controls) and conducted a mega-meta-analysis, working with

more than 11 million SNPs with a very dense coverage of the genomic map [21]. Table 4.1 depicts the list of genes and variants associated with LOAD in this mega-meta-analysis: in addition to the already eight known GWAS-defined genes (*ABCA7*, *BIN1*, *CLU*, *CRI*, *CD2AP*, *EPHA1*, *MS4A4*, and *PICALM*) that have been confirmed (*CD33* gene did not reach here genome-wide significance), 11 new susceptibility loci have been identified in or near plausible candidate genes (*CASS4*, *CELF1*, *FERMT2*, *HLA-DRB5/DRB1*, *INPP5D*, *MEF2C*, *NME8*, *PTK2B*, *SLC24A4/RIN3*, *SORL1*, and *ZCWPWI*). The effects of all these 19 genes on risk for LOAD are exceedingly small (Table 4.1), with allelic ORs between 0.77 (*SORL1*) and 1.22 (*BIN1*); in contrast, the ORs for APOE $\epsilon 4$ are approximately 4 or 15 for one or two $\epsilon 4$ alleles, respectively. That is, one or two copies of the APOE $\epsilon 4$ allele increases the risk for APOE by more than 400 % or 1500 %, whereas one copy of all these non-APOE alleles merely increases or decreases the risk by approximately 30 %, at most. However, the findings from this mega-meta-analysis are, for the most part, not based on the true susceptibility variants but are reflective of their tagging markers, which may harbor greater heterogeneity than the former with respect to alleles and extent of LD. Thus, it remains a possibility that the actual functional susceptibility variants may have bigger effect sizes.

Population Attributable Fraction: Understudied Populations

The cumulative population attributable fraction (e.g., the proportion of LOAD cases in a population that would be prevented if an exposure were eliminated) at each of the 19 non-*APOE* loci identified by the IGAP (Table 4.1) was between 1.1 % (*CASS4* and *SORL1*) and 8.1 % (*BIN1*) and that of APOE was 27.3 % [21]. The remaining genetic risk for LOAD could be due to new common loci, rare variants, structural variants, and gene-gene and gene-environment interactions. Most of large GWAS have identified several variants that affect LOAD susceptibility in non-Hispanic whites of European ancestry. African Americans and other minorities are understudied, and it is unclear whether any of the recently identified loci modify risk of LOAD in racial or ethnic groups other than whites. The ADGC consortium [22] performed a GWAS among the largest sample of African Americans ever assembled for genetic study of LOAD (nearly 2,000 cases and 4,000 cognitively normal elderly controls). The APOE $\epsilon 4$ allele, previously shown to be associated with LOAD in whites, was also implicated in African Americans ($P = 5.5 \times 10^{-47}$, OR = 2.3), and more striking was that the effect size for *ABCA7* was comparable with that observed for APOE. In fact, variants at the *ABCA7* gene increased the risk for LOAD approximately 1.8-fold ($P = 2.21 \times 10^{-9}$) in individuals of African ancestry as opposed to the modest increased risk of 1.15-fold in individuals of European ancestry (Table 4.1). A number of other variants in other genes (*CRI*, *BIN1*, *PICALM*, *CLU*, *EPHA1*, *MS4A* cluster, *CD2AP*, and *CD33*) did not reach the P value cutoff for genome-wide significance in this African American population.

Table 4.1 LOAD-associated GWAS loci in the International Genomics of Alzheimer's Project mega-meta-analysis

SNP	Gene	Mutation type	MAF in controls	PAF (%)	PAF type	OR (95 % CI)	P value	Implicated pathway
rs4147929	<i>ABCA7</i>	Intronic	0.162	2.8	Risk	1.15 (1.11–1.19)	1.1×10^{-15}	A β , cholesterol, immunity
rs6733839	<i>BIN1</i>	Intergenic	0.366	8.1	Risk	1.22 (1.18–1.25)	6.9×10^{-44}	A β , cellular signaling, tau
rs7274581	<i>CASS4</i>	Intronic	0.088	1.1	Preventive	0.88 (0.84–0.92)	2.5×10^{-8}	A β , tau
rs10948363	<i>CD2AP</i>	Intronic	0.255	2.3	Risk	1.10 (1.07–1.13)	5.2×10^{-11}	A β , cellular signaling
rs10838725	<i>CELF1</i>	Intronic	0.312	2.4	Risk	1.08 (1.05–1.11)	1.1×10^{-8}	Cellular signaling
rs9331896	<i>CLU</i>	Intronic	0.398	5.3	Preventive	0.86 (0.84–0.89)	2.8×10^{-25}	A β , cholesterol, immunity
rs6656401	<i>CRI</i>	Intronic	0.191	3.7	Risk	1.18 (1.14–1.22)	5.7×10^{-24}	Immunity
rs11771145	<i>EPHA1</i>	Intergenic	0.350	3.1	Preventive	0.90 (0.88–0.93)	1.1×10^{-13}	Immunity, cellular signaling
rs17125944	<i>FERMT2</i>	Intronic	0.079	1.5	Risk	1.14 (1.09–1.19)	7.9×10^{-9}	Tau
rs9271192	<i>HLA-DRB5</i>	Intergenic	0.277	3.2	Risk	1.11 (1.08–1.15)	2.9×10^{-12}	Immunity
rs35349669	<i>INPP5D</i>	Intronic	0.462	4.6	Risk	1.08 (1.05–1.11)	3.2×10^{-8}	Immunity
rs190982	<i>MEF2C</i>	Intergenic	0.389	2.7	Preventive	0.93 (0.90–0.95)	3.2×10^{-8}	Immunity, cellular signaling
rs983392	<i>MS4A6A</i>	Intergenic	0.406	4.2	Preventive	0.90 (0.87–0.92)	6.1×10^{-16}	Immunity
rs2718058	<i>NME8</i>	Intergenic	0.368	2.9	Preventive	0.93 (0.90–0.95)	4.8×10^{-9}	Cellular signaling
rs10792832	<i>PICALM</i>	Intergenic	0.365	5.3	Preventive	0.87 (0.85–0.89)	9.3×10^{-26}	A β , cellular signaling, tau
rs28834970	<i>PTK2B</i>	Intronic	0.358	3.1	Risk	1.10 (1.08–1.13)	7.4×10^{-14}	Cellular signaling
rs10498633	<i>SLC24A4/RIN3</i>	Intronic	0.212	1.5	Preventive	0.91 (0.88–0.94)	5.5×10^{-9}	Tau
rs11218343	<i>SORL1</i>	Intronic	0.044	1.1	Preventive	0.77 (0.72–0.82)	9.7×10^{-15}	A β , cellular signaling
rs1476679	<i>ZCWPW1</i>	Intronic	0.293	3.2	Preventive	0.91 (0.89–0.94)	5.6×10^{-10}	?

Modified from [21]

LOAD late-onset alzheimer's disease, GWAS genome-wide association study, MAF minor allele frequency, PAF population attributable fraction, OR (95 % CI) odds ratio (95 % confidence interval), A β amyloid-beta

Functional Basis for the LOAD-Associated GWAS Loci

True Functional Variants: Expression Quantitative Trait Loci

The associated SNPs identified through GWAS are unlikely to be functional variants themselves. For any disease-associated SNP, the true variant underlying the phenotype studied may be the GWAS hit itself, a known common SNP in LD with the identified GWAS hit, an unknown common or rare SNP tagged by a haplotype on which the hit occurs, or a linked copy number variant [23]. For all traits studied by GWAS, only 12 % of the associated SNPs are located in, or occur in high LD with, protein-coding regions of genes; the vast majority (80 %) of trait-associated SNPs are located in intergenic regions or noncoding introns [24]. LOAD is not different: taking into account the 19 SNPs reported in the 11 new loci and the 8 previously reported loci associated to LOAD in the IGAP mega-meta-analysis [21], 12 SNPs are located in intronic regions and 7 in intergenic regions (Table 4.1). These findings clearly indicate that follow-up studies should not only examine coding variability but should also pay close attention to the potential roles of these intronic and intergenic regions in the regulation of gene expression. Therefore, GWAS follow-up studies should rely on fine mapping of the associated loci and deep re-sequencing of the associated regions in samples of interest in order to identify all possible functional variants. In addition, it is critical to characterize the novel LOAD candidate variants and genes that are being identified in LOAD-associated GWAS with respect to their influence on gene expression, also known as expression quantitative trait loci (eQTL) studies [25]. The underlying premise of these studies is that the level of the expressed gene transcript (mRNA profiling) from LOAD patients will have changes in comparison to controls, by using data generated from tissue affected by the disease (such as the temporal cortex) or peripheral immune cells [26]. SNPs that influence brain gene expression (eSNPs) constitute an important class of functional variants. In this respect, SNPs in the *CLU* (rs11136000) and *MS4A4A* (rs2304933/rs2304935) genes influenced their expression levels in temporal cortex [27]: the LOAD-protective *CLU* and the risky *MS4A4A* alleles both occurred in conjunction with elevated levels of brain expression, implicating regulatory genetic variation for these genes in LOAD risk. In a systematic examination of *CLU*, *CRI*, *ABCA7*, *BIN1*, *PICALM*, and *MS4A6A/MS4A6E* loci for LOAD, coding variability might explain only the *ABCA7* association with LOAD, but common coding variability did not explain any of the other loci; in addition, none of these loci had eQTL effects and the regional expression of each of the loci did not match the pattern of brain regional distribution in Alzheimer pathology [23].

Pathogenic Pathways Implicated in LOAD from GWAS Loci

The LOAD candidate genes make biological sense and have identified different pathways involved in LOAD pathogenesis [4, 21, 28]. As suggested by Table 4.1, the implicated pathways are:

- A/Amyloid-beta metabolism (production, degradation, and clearance): *APOE*, *CLU*, *ABCA7*, *PICALM*, *BINI*, *CD2AP*, *SORL1*, *CASS4*, and *CD33* [29]
- B/Immune system function (both innate and adaptive): *CLU*, *CRI*, *ABCA7*, *MS4A* cluster, *CD33*, *EPHA1*, *HLA-DRB5/DRB1*, *INPP5D*, and *MEF2C*
- C/Cholesterol metabolism: *APOE*, *CLU*, and *ABCA7*
- D/Synaptic cell functioning mechanisms and cell membrane processes (endocytosis): *PICALM*, *BINI*, *CD33*, *CD2AP*, *EPHA1*, *SORL1*, *CELF1*, *NME8*, *MEF2C*, and *PTK2B*
- E/Tau pathology (microtubule stability, tau phosphorylation/aggregation, and neurofibrillary tangle formation): *CASS4*, *FERMT2*, *SLC24A4/RIN3*, *BINI* [30], and *PICALM* [31]

Exactly how *APOE* might cause LOAD is a matter of debate, and as well as being the main transporter of cholesterol and other lipids into the brain, it is also thought to remove amyloid-beta from the brain. Ultimately, the validation of the pathogenic mechanisms of all these LOAD GWAS loci will require comprehensive functional studies in in vitro systems, in vivo animal models, and clinical samples.

Examining Genetic Influences on Endophenotypes

Endophenotypes are biologically relevant, quantitative, and heritable phenotypes. There are many endophenotypes that are currently utilized or are excellent candidates for genetic studies of LOAD, including cerebrospinal fluid measures of amyloid-beta, tau and phosphorylated tau, neuroimaging measures in magnetic resonance imaging (MRI) and positron emission tomography (PET) scans (such as hippocampal volume), and cognitive measures [25]. Genetic studies of LOAD endophenotypes are an effective approach for identifying disease risk loci that are complementary to case-control association studies, and these genetic variants might be implicated not only with risk for LOAD but also with age at onset or rate of progression. Cognitive endophenotypes (e.g., level of cognitive function and rate of decline in cognition) can help to detect genetic risk factors attributable to the preclinical and subclinical change in cognition in LOAD. For example, the simultaneous consideration of the joint effects of eight non-*APOE* LOAD-associated GWAS loci (*ABCA7*, *BINI*, *CD2AP*, *CLU*, *CRI*, *MS4A4E*, *MS4A6A*, and *PICALM*) aggregated as a cumulative genetic risk score predicts accelerated progression from mild cognitive impairment (MCI) to LOAD in those subjects with higher scores [32]. Moreover, MCI patients with the *APOE* $\epsilon 4$ allele are more likely to convert to LOAD as compared to those without the *APOE* $\epsilon 4$ allele [33]. No clear profile has emerged from studies of the relation between genotype and amyloid or tau phenotype in cerebrospinal fluid: whereas no evidence for association between variants in *BINI*, *CLU*, *CRI*, and *PICALM* genes and amyloid-beta and phosphorylated tau levels in cerebrospinal fluid has been found in a study [34], *APOE* $\epsilon 4$ allele, *CLU*, and *MS4A4A* genetic variants were associated with significantly reduced amyloid-beta levels in cerebrospinal fluid in other study [35]. Investigating whether LOAD-associated GWAS loci influence MRI measures

(hippocampal and amygdala volumes and entorhinal cortex and temporal pole cortex thicknesses), the *APOE* $\epsilon 4$ allele and *PICALM* and *CRI* genotypes have been significantly associated with these neuroimaging measures [36].

The Whole Exome and Whole Genome Sequencing Approach

Common Versus Rare Variants

A proportion of heritability (the portion of phenotypic variance in a population attributable to additive genetic factors) is apparently unexplained by GWAS findings. Explanations for this “missing heritability” include rarer variants (possibly with larger effects) that are poorly detected by available genotyping arrays that focus on variants present in 5 % or more of the population; structured variants poorly captured by existing arrays, including copy number variants such as insertions and deletions and copy neutral variation such as inversions and translocations; low power to detect gene-gene interactions; and inadequate accounting for shared environment among relatives [37]. It is likely that a substantial portion of the genetic risk underlying LOAD is actually conferred by rare sequence variants, those occurring with a frequency <1 % in the general population, and possibly of relatively large genetic effect (e.g., with odds ratios >2). Rare variants are much more likely to have functional consequence than the more common variants; in fact, regulatory regions show preferential exclusion of common variants relative to rare ones just like protein-coding sequence [38]. GWAS are by design powered to detect association with causal variants that are relatively common in the population, and current microarray technology is not designed for de novo identification of rare sequence variants. Thus, the identification of presumed disease-associated rare variants requires deep re-sequencing in suitable data sets, either small scale (e.g., previously associated GWAS regions) or large-scale (whole exome or whole genome). Whole exome sequencing is most often chosen for monogenic Mendelian diseases, largely because of its low cost compared with whole genome sequencing (the exome is 1–2 % of the whole genome) and the notion that most sequence variations leading to a severe phenotypic effect are located in the coding part of the genome [4]. Whole exome sequencing is capable of identifying not only very rare Mendelian causes of disease but also low-frequency variability with medium-effect sizes modulating disease development. A significant proportion of EOAD is caused by autosomal dominant, fully penetrant mutations. LOAD recurs within families more often than expected by chance alone, and this observed familial recurrence could be attributed to genetic loci with large phenotypic effects and reduced penetrance (possibly recessive loci) [10]. With monogenic recessive contributions to LOAD, 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. So far, the role of recessive mutations in LOAD has been considerably overlooked.

Rare Monogenic Forms of LOAD

TREM2

Homozygous loss-of-function mutations in *TREM2* gene, encoding the triggering receptor expressed on myeloid cells 2 protein, have previously been associated with an autosomal recessive form of early-onset dementia presenting with bone cysts and consequent fractures called polycystic lipomembranous osteodysplasia with sclerosing leukoencephalopathy or Nasu-Hakola disease. Homozygous *TREM2* mutations have been recently identified in three Turkish patients presenting with a clinical phenotype associated with frontotemporal dementia and with leukodystrophy but without any bone-associated symptoms [39]. Whereas severe and early-onset disease is caused by homozygous loss-of-function mutations in *TREM2*, heterozygous loss-of-function variants are associated with LOAD. For example, in 2,000 LOAD patients and 4,000 controls, a rare missense mutation (rs75932628-T) in *TREM2*, which was predicted to result in a R47H substitution, showed a strong, highly significant association with LOAD ($P=9.0\times 10^{-9}$, OR=5.05), with a minor allele frequency among healthy controls of 0.12 % in the United States [40]. Similarly, in 2,261 Icelandic participants, the T allele of rs75932628 in *TREM2* was found to confer a significant risk of LOAD ($P=3.42\times 10^{-10}$, OR=2.9), with a minor allele frequency of 0.63 % in healthy controls [41]. Consequently, this R47H variant of *TREM2* is a low-prevalence variant that increases LOAD risk with a moderate-to-high effect size, similar to that of the *APOE* $\epsilon 4$ allele. Neurodegeneration in *TREM2-associated* LOAD is probably driven by a chronic inflammatory process with dysfunction in the microglial phagocytosis [42].

APP

About 25 coding mutations in the *APP* gene have resulted in EOAD, but until now, mutations in *APP* had not been implicated in LOAD. In a set of whole genome sequence data from 1,795 Icelanders, the A allele of rs63750847 results in an alanine to threonine substitution at position 673 in *APP* (*A673T*) and was found to be significantly more common in the elderly control group aged 85 or greater than in the LOAD group (0.62 % versus 0.13 %, $P=4.78\times 10^{-7}$, OR=0.189) [43]. In addition, the cognitive function of elderly noncarriers remained poorer than for carriers of *A673T* after removing LOAD cases. *A673T* represents the first example of a rare sequence variant conferring strong protection against LOAD and also protecting against cognitive decline in the elderly without LOAD. The *A673T* substitution is critical for reducing the production of amyloid-beta. The complete absence of the *A673T* variant in a large cohort of Asian subjects [44] suggests that this is possibly an ethnicity-specific variant.

MAPT

In a combined analysis of 15,369 subjects, re-sequencing at the gene encoding for the microtubule associated protein tau (*MAPT*) discovered that the rare substitution *A152T* within exon 7 of *MAPT* increases the risk for LOAD (0.69 % in patients versus 0.30 % in controls, $P=4.0\times 10^{-3}$, OR=2.3) and also for frontotemporal dementia (0.89 % in patients, $P=5.0\times 10^{-4}$, OR=3.0) [45]. This study emphasizes the point that statistical evaluation of the role of rare sequence variants poses a challenge, and no thresholds for rare variant significance have been established. The functional studies show that the *A152T* in *MAPT* causes a pronounced decrease in microtubule stability and might enhance the level of tau oligomers. This is another example that rare variants can increase the risk for complex diseases with heterogeneous phenotypes.

FRMD4A

In a meta-analysis of EADI and GERARD consortia and a combined analysis of five additional case-control studies (10,000 LOAD and 16,000 controls), the AAC haplotype in the *FRMD4A* locus was associated with increased LOAD risk ($P=1.1\times 10^{-10}$, OR=1.68) when compared with most frequent *GGT* haplotype [46]. As the AAC haplotype is rare (with a mean frequency of 2 % in Caucasian populations), this might explain why the locus was not detected in previous GWAS based on single-SNPs analyses, as SNPs with low frequency are poorly imputed even when using the 1000 Genomes data set. Therefore, other complementary approach to GWAS is this example of genome-wide haplotype association study. The protein encoded by *FRMD4A* is involved in cell structure, transport, and signaling functions.

Gene-Gene Interactions (Epistasis)

Evidence is accumulating that a pronounced part of the still elusive genetic variability in complex diseases could be due to ignored epistatic effects [47]. The term epistasis is conventionally used when an increased risk is only seen in the presence of two genetic factors and not seen when they act apart. In such cases, studies that examine simple loci individually, such as most GWAS, will fail to detect an effect. To understand the causes of LOAD, one needs to study not simple factors one at a time but interactions between genetic risk factors. In the case of LOAD, epistasis is likely to play a major part, in view of the high heritability of the disease. Epistasis had previously proved hard to demonstrate, mainly because sample sets had been too small and poorly characterized and inappropriate statistical methods had been used. The Epistasis Project [48] was designed to avoid these problems, with a multinational collaboration of 7 LOAD research groups from the UK, Spain, the

Netherlands, and Germany, contributing DNA samples from 1,757 LOAD cases and 6,295 controls. A typical GWAS may examine perhaps 500,000 loci, but the number of potential two-way interactions between these 500,000 loci is >100 billion (10^{11}). In order therefore to reduce the number of potential interactions to a manageable figure, a hypothesis-driven approach is required, and consequently, a selection of gene-gene interactions should be chosen according to prior evidence of a statistical interaction and a plausible biological hypothesis [49]. The chosen interactions in the Epistasis Project were involved in various pathogenic networks that contribute to the development of LOAD (lipid metabolism, amyloid-beta metabolism, inflammation, oxidative stress, and insulin metabolism), and the “synergy factor” [50] (equivalent to the interaction term defined by two binary factors in a logistic regression model) was used to measure the gene-gene interaction. In the inflammation pathway, the Epistasis Project has demonstrated that the interaction between the interleukin-6 proinflammatory cytokine and the interleukin-10 anti-inflammatory cytokine genes [48] and the interaction between the aromatase (a rate-limiting enzyme in the synthesis of estrogens) and the interleukin-10 genes [51] are both associated with increased LOAD risk. In the oxidative stress pathway, the Epistasis Project has revealed an increased LOAD risk due to the interaction between the hemochromatosis and transferrin genes [52] and the interaction between the glutathione S-transferase and the gene cluster of the hematopoietically expressed homeobox, the insulin-degrading enzyme, and the kinesin family member 11 [53]. In the future, to achieve higher power for such gene-gene interaction studies, larger sample sizes are needed, such as that of the IGAP mega-meta-analysis of GWAS [21].

References

1. Ertekin-Taner N. Genetics of Alzheimer disease in the pre- and post-GWAS era. *Alzheimers Res Ther.* 2010;2:3.
2. Bertram L. Alzheimer's genetics in the GWAS era: a continuing story of 'replications and refutations'. *Curr Neurol Neurosci Rep.* 2011;11:246–53.
3. Tanzi RE. The genetics of Alzheimer disease. *Cold Spring Harb Perspect Med.* 2012;2:a006296.
4. Bettens K, Sleegers K, Van Broeckhoven C. Genetic insights in Alzheimer's disease. *Lancet Neurol.* 2013;12:92–104.
5. 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;63:168–74.
6. Gerrish A, Russo G, Richards A, Moskvina V, Ivanov D, Harold D, et al. The role of variation at AβPP, PSEN1, PSEN2, and MAPT in late onset alzheimer's disease. *J Alzheimers Dis.* 2012;28:377–87.
7. Strittmatter WJ, Saunders AM, Schmechel D, Pericak-Vance M, Enghild J, Salvesen GS, Roses AD. Apolipoprotein E: high-avidity binding to beta-amyloid and increased frequency of type 4 allele in late-onset familial Alzheimer disease. *Proc Natl Acad Sci U S A.* 1993;90:1977–81.
8. Bertram L, Tanzi RE. The genetics of Alzheimer's disease. In: Teplow D, editor. *Progress in molecular biology and translational science*, vol. 107. Burlington: Academic; 2012. p. 79–100.
9. 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; 16:903–7.

10. Guerreiro RJ, Gustafson DR, Hardy J. The genetic architecture of Alzheimer's disease: beyond APP, PSENs and APOE. *Neurobiol Aging*. 2012;33:437–56.
11. 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.
12. International HapMap Consortium. A genome haplotype map of the human genome. *Nature*. 2005;437:1299–320.
13. International HapMap Consortium, Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, et al. A second generation human haplotype map of over 3.1 million SNPs. *Nature*. 2007;449:851–61.
14. Schellenberg GD, Montine TJ. The genetics and neuropathology of Alzheimer's disease. *Acta Neuropathol*. 2012;124:305–23.
15. 1000 Genomes Project Consortium, Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, et al. A map of human genome variation from population-scale sequencing. *Nature*. 2010;467:1061–73.
16. 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;41:1088–93.
17. Lambert JC, Heath S, Even G, Campion D, Sleegers K, Hiltunen M, et al. Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. *Nat Genet*. 2009;41:1094–9.
18. Seshadri S, Fitzpatrick AL, Ikram MA, DeStefano AL, Gudnason V, Boada M, et al. Genome-wide analysis of genetic loci associated with Alzheimer disease. *JAMA*. 2010;303:1832–40.
19. 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;43:429–35.
20. Naj AC, Jun G, Beecham GW, Wang LS, Vardarajan BN, Buross J, et al. Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. *Nat Genet*. 2011;43:436–41. doi:[10.1038/ng.2802](https://doi.org/10.1038/ng.2802).
21. 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;45:1452–8.
22. Reitz C, Jun G, Naj A, Rajbhandary R, Vardarajan BN, Wang LS, et al. Variants in the ATP-binding cassette transporter (ABCA7), apolipoprotein E ϵ 4, and the risk of late-onset Alzheimer disease in African Americans. *JAMA*. 2013;309:1483–92.
23. Holton P, Ryten M, Nalls M, Trabzuni D, Weale ME, Hernandez D, et al. Initial assessment of the pathogenic mechanisms of the recently identified Alzheimer risk loci. *Ann Hum Genet*. 2013;77:85–105.
24. Manolio TA. Genome wide association studies and assessment of the risk of disease. *N Engl J Med*. 2010;363:166–76.
25. Ertekin-Taner N, De Jager PL, Yu L, Bennett DA. Alternative approaches in gene discovery and characterization in Alzheimer's disease. *Curr Genet Med Rep*. 2013;1:39–51.
26. Stranger BE, Stahl EA, Raj T. Progress and promise of genome-wide association studies for human complex trait genetics. *Genetics*. 2011;187:367–83.
27. Allen M, Zou F, Chai HS, Younkin CS, Crook J, Pankratz VS, et al. Novel late-onset Alzheimer disease loci variants associate with brain gene expression. *Neurology*. 2012;79:221–8.
28. Morgan K. The three new pathways leading to Alzheimer's disease. *Neuropathol Appl Neurobiol*. 2011;37:353–7.
29. Griciuc A, Serrano-Pozo A, Parrado AR, Lesinski AN, Asselin CN, Mullin K, et al. Alzheimer's disease risk gene CD33 inhibits microglial uptake of amyloid beta. *Neuron*. 2013;78:631–43.
30. Chapuis J, Hansmannel F, Gistelincq M, Mounier A, Van Cauwenberghe C, Kolen KV, et al. Increased expression of BIN1 mediates Alzheimer genetic risk by modulating tau pathology. *Mol Psychiatry*. 2013;18:1225–34. doi:[10.1038/mp.2013.1](https://doi.org/10.1038/mp.2013.1).

31. Ando K, Brion JP, Stygelbout V, Suain V, Authelet M, Dedecker R, et al. Clathrin adaptor CALM/PICALM is associated with neurofibrillary tangles and is cleaved in Alzheimer's brains. *Acta Neuropathol.* 2013;125:861–78.
32. Rodríguez-Rodríguez E, Sánchez-Juan P, Vázquez-Higuera JL, Mateo I, Pozueta A, Berciano J, et al. Genetic risk score predicting accelerated progression from mild cognitive impairment to Alzheimer's disease. *J Neural Transm.* 2013;120:807–12.
33. Elias-Sonnenschein LS, Viechtbauer W, Ramakers IH, Verhey FR, Visser PJ. Predictive value of APOE-ε4 allele for progression from MCI to AD-type dementia: a meta-analysis. *J Neurol Neurosurg Psychiatry.* 2011;82:1149–56.
34. Kauwe JS, Cruchaga C, Karch CM, Sadler B, Lee M, Mayo K, et al. Fine mapping of genetic variants in BIN1, CLU, CR1 and PICALM for association with cerebrospinal fluid biomarkers for Alzheimer's disease. *PLoS One.* 2011;6:e15918.
35. Elias-Sonnenschein LS, Helisalmi S, Natunen T, Hall A, Paajanen T, Herukka SK, et al. Genetic loci associated with Alzheimer's disease and cerebrospinal fluid biomarkers in a Finnish case–control cohort. *PLoS One.* 2013;8:e59676.
36. Biffi A, Anderson CD, Desikan RS, Sabuncu M, Cortellini L, Schmansky N, et al. Alzheimer's Disease Neuroimaging Initiative (ADNI). Genetic variation and neuroimaging measures in Alzheimer disease. *Arch Neurol.* 2010;67:677–85.
37. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. *Nature.* 2009;461:747–53.
38. Zhu Q, Ge D, Maia JM, Zhu M, Petrovski S, Dickson SP, et al. A genome-wide comparison of the functional properties of rare and common genetic variants in humans. *Am J Hum Genet.* 2011;88:458–68.
39. Guerreiro RJ, Lohmann E, Brás JM, Gibbs JR, Rohrer JD, Gurunlian N, et al. Using exome sequencing to reveal mutations in TREM2 presenting as a frontotemporal dementia-like syndrome without bone involvement. *JAMA Neurol.* 2013;70:78–84.
40. Guerreiro R, Wojtas A, Bras J, Carrasquillo M, Rogava E, Majounie E, Alzheimer Genetic Analysis Group, et al. TREM2 variants in Alzheimer's disease. *N Engl J Med.* 2013;368:117–27.
41. 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;368:107–16.
42. Neumann H, Daly MJ. Variant TREM2 as risk factor for Alzheimer's disease. *N Engl J Med.* 2013;368:182–4.
43. 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;488:96–9.
44. Seng Ting SK, Chong MS, Kandiah N, Hameed S, Tan L, Au WL, et al. Absence of A673T amyloid-β precursor protein variant in Alzheimer's disease and other neurological diseases. *Neurobiol Aging.* 2013;34:2441e7–8.
45. Coppola G, Chinnathambi S, Lee JJ, Dombroski BA, Baker MC, Soto-Ortolaza AI, et al. Evidence for a role of the rare p.A152T variant in MAPT in increasing the risk for FTD-spectrum and Alzheimer's diseases. *Hum Mol Genet.* 2012;21:3500–12.
46. Lambert JC, Grenier-Boley B, Harold D, Zelenika D, Chouraki V, Kamatani Y, et al. Genome-wide haplotype association study identifies the FRMD4A gene as a risk locus for Alzheimer's disease. *Mol Psychiatry.* 2013;18:461–70.
47. Zuk O, Hechter E, Sunyaev SR, Lander ES. The mystery of missing heritability: genetic interactions create phantom heritability. *Proc Natl Acad Sci U S A.* 2012;109:1193–8.
48. Combarros O, van Duijn CM, Hammond N, Belbin O, Arias-Vásquez A, Cortina-Borja M, et al. Replication by the Epistasis Project of the interaction between the genes for IL-6 and IL-10 in the risk of Alzheimer's disease. *J Neuroinflammation.* 2009;6:22.
49. Combarros O, Cortina-Borja M, Smith AD, Lehmann DJ. Epistasis in sporadic Alzheimer's disease. *Neurobiol Aging.* 2009;30:1333–49.

50. Cortina-Borja M, Smith AD, Combarros O, Lehmann DJ. The synergy factor: a statistic to measure interactions in complex diseases. *BMC Res Notes*. 2009;2:105.
51. Medway C, Combarros O, Cortina-Borja M, Butler HT, Ibrahim-Verbaas CA, de Bruijn RF, et al. The sex-specific associations of the aromatase gene with Alzheimer's disease and its interaction with IL10 in the Epistasis Project. *Eur J Hum Genet*. 2014;22(2):216–20. doi:[10.1038/ejhg.2013.116](https://doi.org/10.1038/ejhg.2013.116).
52. Lehmann DJ, Schuur M, Warden DR, Hammond N, Belbin O, Kölsch H, et al. Transferrin and HFE genes interact in Alzheimer's disease risk: the Epistasis Project. *Neurobiol Aging*. 2012;33:202.e1–13.
53. Bullock JM, Medway C, Cortina-Borja M, Turton JC, Prince JA, Ibrahim-Verbaas CA, et al. Discovery by the Epistasis Project of an epistatic interaction between the GSTM3 gene and the HHEX/IDE/KIF11 locus in the risk of Alzheimer's disease. *Neurobiol Aging*. 2013;34:1309.e1–7.