Power and Pitfalls of the Genome-Wide Association Study Approach to Identify Genes for Alzheimer's Disease

Richard Sherva · Lindsay A. Farrer

Abstract Until recently, the search for genes contributing to Alzheimer's disease (AD) had been slow and disappointing, with the notable exception of the *APOE* ε4 allele, which increases risk and reduces the age at onset of AD in a dose-dependent fashion. Findings from genome-wide association studies (GWAS) made up of fewer than several thousand cases and controls each have not been replicated. Efforts of several consortia—each assembling much larger datasets with sufficient power to detect loci conferring small changes in AD risk—have resulted in robust associations with many novel genes involved in multiple biological pathways. Complex data mining strategies are being used to identify additional members of these pathways and gene–gene interactions contributing to AD risk. Guided by GWAS results, next-generation sequencing and functional studies are under way with the hope of helping us better understand AD pathology and providing new drug targets.

Keywords Alzheimer's disease · Genome-wide association study · Apolipoprotein E · Alzheimer's disease susceptibility genes

R. Sherva
Department of Medicine (Biomedical Genetics),
Boston University School of Medicine,
Boston, MA 02118, USA

L. A. Farrer (✉)
Departments of Medicine (Biomedical Genetics), Neurology,
Ophthalmology, Genetics and Genomics, Biostatistics,
and Epidemiology, Boston University Schools of Medicine and
Public Health, Biomedical Genetics L-320,
72 East Concord Street,
Boston, MA 02118, USA
e-mail: farrer@bu.edu

Introduction

Early genetic studies of Alzheimer's disease (AD) focused primarily on families displaying autosomal dominant inheritance of the disorder and used linkage-based tests to identify genetic regions harboring disease-causing mutations. These efforts pinpointed discrete regions on chromosomes 1 [1], 14 [2, 3], and 21 [4] and facilitated the eventual identification of mutations in the culprit genes in each of these regions, specifically those encoding presenilin 2 (*PS2*) [5], presenilin 1 (*PS1*) [6], and amyloid precursor protein (*APP*) [7], respectively. Collectively, mutations in these genes account for fewer than 1% of AD cases in the population and are associated with onset of disease symptoms generally between ages 30 and 60 years [8]. Despite their rarity, subsequent studies of these mutations in cultured cells and animal models greatly increased our understanding of AD pathogenesis [9]. Successful identification of genes for early-onset AD emboldened several research teams to use linkage approaches to search for genes causing the more common late-onset form of AD. These studies were conducted before high-density maps of single nucleotide polymorphisms (SNPs) were available and relied on relatively sparse microsatellite marker coverage across the genome. Although evidence for linkage was obtained at many locations throughout the genome, the most consistently replicated regions were on chromosomes 6, 9, 10, 12, and 19 [9]. The linkage peaks were generally very broad and contained biologically plausible candidate genes with supporting evidence from candidate gene association studies, including *APOE, A2M, LRP1, GAPD, IDE, PLAU, CTNNA3, GTSO1/GTSO2,* and *UBQLN* [9, 10]. Disappointingly, *APOE* is the only gene among this group that has been clearly established as a genetic risk factor for late-onset AD, and it should be noted that linkage studies—rather than a

candidate gene approach—led to its association with AD [11]. These linkage studies were conducted in a wide range of ethnic groups and were subject to the limitations of linkage analysis to varying degrees, including low resolution, locus heterogeneity, and insufficient sample size. Given these caveats, it is not surprising that subsequent fine-mapping and candidate gene studies largely failed to identify the causal variants contributing to the linkage peaks.

Candidate Gene Studies

As of September 2010, the Alzheimer Disease Forum in the AlzGene database [12] (http://www.alzforum.org/res/com/gen/alzgene) had curated information from 1,375 candidate gene and genome-wide association studies (GWAS) about 664 candidate genes. Among the genes implicated by hypothesis-driven approaches, none of the associations have been consistently replicated (as discussed above, *APOE* was a positional candidate). Several reports summarized by Ertekin-Taner [10] indicate possible reasons for the lack of replication of genetic results across studies, including 1) initial false-positive results followed by lack of replication; 2) lack of adequate power to detect or replicate an association result because of sample size (false negatives); 3) lack of informative markers at the locus; 4) locus heterogeneity (different genes underlying the same AD phenotype); and 5) clinical heterogeneity (multiple clinical subtypes associated with different sets of susceptibility genes). Lack of consistency across studies may also be attributed to intralocus (ie, allelic) heterogeneity. An example of this phenomenon is *SORL1*, for which there is compelling evidence for two discrete, biologically relevant variants, even in the same population sample [13]. Incidentally, at the time of this writing, *SORL1* is the top-ranked gene identified as a biological candidate in the AlzGene database.

One of the limitations of gene association studies—which may explain in part the disparate findings across studies—is that they are very sensitive to allele frequencies. Results may be biased if the control group has allele frequencies that are not representative of the population from which the patient population is drawn. Spurious associations that are related to some extraneous factor unrelated to disease are more likely to be detected in unrelated case-control samples than in studies of families [14]. These biases can be minimized by studying gene associations within families. Second, it is likely that genes other than *APOE* will not individually account for much genetic variance of disease risk; therefore, some may be difficult to detect and consistently replicate in outbred populations. The most robust associations likely will be detected in multiple ethnically diverse populations, as exemplified by findings for *APOE* [8], *SORL1* [13], and *ACE* [15, 16].

Genome-Wide Association Studies

The advent of affordable, commercially available micro-array chips containing up to 1.2 million SNPs heralded the GWAS era. This technology overcame several limitations of linkage and candidate gene studies. First, precision of gene searches was greatly increased. Second, SNP panels used in GWAS potentially contain the actual causal variants, whereas non–gene-based microsatellite marker panels used for linkage studies generally do not. Third, the GWAS approach is agnostic with respect to biological mechanisms underlying the disease and hence has an advantage over candidate gene studies, the success of which depends on correct knowledge about the pathways leading to disease. There are many instances of robust associations identified by GWAS with genes that were non-obvious candidates based on contemporary knowledge of the disease pathophysiology [17, 18]. Fourth, an investigator may correctly hypothesize the genetic mechanism for the disease but fail to detect association using a candidate gene approach if the risk variant is outside the coding or presumed regulatory regions, as was exemplified by a GWAS for prostate cancer in which several robustly associated and potentially functional SNPs are located in a "gene desert" between two oncogenes, *FAM84B* and c-*MYC* [19].

The advantage afforded by the huge number of markers tested in GWAS is also the source of the primary shortfall of this approach: there will almost always be apparently "significant" findings when several hundred thousand or more tests are performed, and nearly all of them will be false-positive results. Generally, researchers attempt to limit the number of false positives by setting stringent P value thresholds for significance (generally 5×10^{-8} in Caucasians), although there are several examples of associations exceeding this threshold that were never subsequently replicated, including many in the AD GWAS discussed below. Unfortunately, this stringent significance threshold greatly increases the probability of false negatives.

Since 2007, there have been 10 published GWAS of AD performed by individual research teams reporting on nine distinct datasets. A broad range of sample sizes and ethnic groups is represented in these studies [20–29], which are summarized in Table 1. Although all but one of these studies confirm the *APOE* region signal on chromosome 19, collectively, few novel AD risk genes were identified, and nearly all these findings were not confirmed in independent samples. The first AD GWAS evaluated association with a panel of about 17,000 coding SNPs and used a two-stage genotyping protocol [20]. The first stage analyzed SNPs in two pools of DNA from AD cases and controls, respectively, and SNPs that showed the greatest difference in pooled allele frequency between cases versus controls were genotyped in subjects individually in the

Table 1 Summary of Alzheimer's disease genome-wide association studies

Study	Family/case-control	Recruitment site(s)	SNPs, n (platform)	Cases, n (D/R)	Controls, n (D/R)	Genes identified	Smallest non-APOE P value	Notes
Grupe et al. [20]	Case-control	United States and United Kingdom	17,343 (rtPCR)	380/1428	396/1666	APOE, ACAN, BCR CTSS, EBF3, FAM63A, GALP GWA_14q32.13 GWA_7p15.2 LMNA, LOC651924, MYH13, PCK1, PGBD1, TNK1, TRAK2, UBD	6×10^{-5}	Pooled genotyping
Coon et al. [22]; Reiman et al. [28]	Case-control	United States, The Netherlands	502,627 (Affymetrix 500K[a])	446/415	290/260	APOE, GAB2	9×10^{-11}	Histopathological AD diagnosis, functional data
Li et al. [23]	Case-control	Canada and United Kingdom	469,438 (Affymetrix 500K[a])	753/418	736/249	APOE, APOC1, GOLM1, GWA_15q21.2,GWA_9p24.3	8×10^{-5}	Tested for age at onset of effects
Poduslo et al. [27]	Family based and case-control	United States	489,218 (Affymetrix 500K[a])	9/199	10/225	TRPC4AP	6×10^{-11}	Two extended pedigrees used for discovery, sequencing
Abraham et al. [24]	Case-control	United Kingdom	561,494 (Illumina HumanHap 300, Illumina Sentrix HumanHap 240S[b])	1082/0	1239/1400	APOE, LRAT	6×10^{-7}	Pooled genotyping
Bertram et al. [21]	Family based	United States	484,522 (Affymetrix 500K[a])	941/1767	404/838	APOE, ATXN1, CD33, GWA_14q31	4×10^{-6}	P values from less powerful FBAT
Beecham et al. [26]	Case-control	United States	532,000 (Illumina 550 BeadChip[b])	492/238	496/220	APOE, FAM113B, MTHFD1L	1×10^{-6}	Nominal association with several AD candidates
Carrasquillo et al. [25]	Case-control	United States	313,504 (Illumina HumanHap 300-Duo[b])	844/1547	1255/1209	APOE, PCDH11X	4×10^{-12}	PCDH11X is on chromosome X
Sherva et al. [29]	Case-control	Israel	286,316 (Illumina HumanHap 300[b])	124/142	5509/7254	AGER, AGPAT1, APBA1, ATP6V0A4, TMEM132C	2.7×10^{-7}	Isolated Israeli-Arab population

[a] Santa Clara, CA

[b] San Diego, CA

AD Alzheimer's disease, D/R discovery sample/replication sample, FBAT family-based association test, rtPCR reverse transcription polymerase chain reaction, single nucleotide polymorphism

second stage. Except for SNPs in the association peak near *APOE*, none of the results remained significant after follow-up, perhaps due to the sample size, poor sensitivity of some assays for allelotyping in pooled DNA samples, or the less stringent significance threshold used to identify SNPs for follow-up. This method can greatly reduce the cost of genotyping but requires careful standardization of the DNA concentration in each pool in order to avoid bias. Coon et al. [22] were also unable to detect novel genome-wide significant findings in their GWAS of a cohort using a much denser panel of SNPs. A subsequent analysis of this dataset revealed significant association with haplotypes in *GAB2* among individuals who had one or more *APOE* ε4 allele [28]. *GAB2* is the principal activator of the phosphatidylinositol-3-kinase signaling pathway in T cells, which activates a pathway that ultimately suppresses tau phosphorylation [30]. Although these results require replication before they can be firmly accepted, they are particularly interesting because they are in accord with evidence from neurobiological studies implicating tau protein-related processes in AD. Also, it was shown recently that the protective haplotype identified previously is associated with higher glucose metabolism in areas of the brain affected by AD in cognitively normal *APOE* ε4 homozygotes [31].

Abraham et al. [24], using a two-stage genotyping protocol, identified an SNP in *LRAT* after individual genotyping with near genome-wide significance, but this result has not been replicated in subsequent studies. Bertram et al. [21] applied family-based association tests and also jointly tested association of markers with the risk and age at onset of AD in a dataset of small nuclear families. Genome-wide significant associations were identified in *APOC1* adjacent to *APOE* and *ATXN1*, two predicted genes on chromosomes 14 and 19, and a noncoding RNA on chromosome 18. The SNPs on chromosomes 14 and 19 showed evidence for association in their replication sample, with the former driven by age at onset and the latter with similar evidence from age at onset and presence of AD. The most noteworthy results in the Li et al. [23] study were the findings for three SNPs among the 120 most significant associations in their discovery set that were replicated in an independent sample; however, none of these results attained genome-wide significance. Two of these SNPs are located on chromosome 9, one in the gene encoding *GOLPH2*, which mediates protein transport through the Golgi apparatus, and one in an intergenic region near a noncoding RNA. The third SNP is located on chromosome 6 between genes encoding an adenosine triphosphate–ase (*ATP8B4*) and a fatty acid transporter (*SLC27A2*).

Carrasquillo et al. [25] identified genome-wide significant association with an SNP in protocadherin 11, X-linked (*PCDH11X),* a member of a sex-specific family of calcium-dependent cell adhesion and recognition proteins that function in central nervous system development. This finding has not yet been confirmed. Inexplicably, the association was significant in females in a dose-dependent manner but not in hemizygous males. Beecham et al. [26] did not observe any Bonferroni-adjusted, genome-wide significant results outside the *APOE* region in a relatively small discovery sample of slightly less than 500 cases and 500 controls. However, they highlighted a result for a vitamin D receptor SNP that was significant according to a relatively liberal false-discovery rate threshold of 0.17 and was also nominally significant in an independent replication sample. In a subsequent study, these investigators obtained a promising result with an intronic SNP in *MTHFD1L*, which was subsequently boosted to a genome-wide significant level with an enlarged dataset of more than 900 cases and 1100 controls [32]. Poduslo et al. [27]compared data for more than 500,000 SNPs from nine AD cases in two extended families with 70 controls, including 10 unaffected family members and 60 unrelated individuals in the Centre Etude Polymorphism Humain (CEPH) reference panel. They observed association ($P \leq 3.1 \times 10^{-9}$) with six SNPs in *TPRC4*; however, the significance of these findings is greatly inflated because the analysis ignored the familial relationships among the AD subjects and potential bias due to genotyping differences between the AD cases and CEPH controls.

A recently published AD GWAS used a unique sample of AD cases and controls from a genetically isolated and highly inbred population of Israeli-Arabs with has an unusually high prevalence of AD despite having a lower frequency of the *APOE* ε4 alleles than other Caucasian populations [33]. To overcome the limitations inherent to the small sample size (124 cases and 142 controls), Sherva et al. [29] focused the initial association tests on regions containing long stretches of homozygosity and, hence, recessive AD susceptibility loci that are presumably rare in outbred populations. As expected, a relatively high degree of inbreeding was observed. However, there was also evidence of extensive admixture and population stratification, indicating that the population is not as isolated as was presumed. Even more surprising, the excess homozygosity was significantly greater in controls. Although none of the SNPs in regions of homozygosity were significantly associated with AD after multiple test correction, several interesting candidate genes were identified in these regions, including *AGER*, *AGPAT1*, and *APBA1*. Analysis of the genome-wide data revealed AD associations with SNPs in three other genes (*ATP6V0A4*, *TMEM132C,* and *GLOD4*) and an intergenic region between *RGS6* and *DPF3* (at $P < 10^{-5}$), and these findings were nominally replicated in several other GWAS datasets. Nonetheless, associations with these loci do not explain the increased disease prevalence in this population.

In general, the results from this group of GWAS were very discouraging and questioned not only the utility of this approach in AD but also the assumption that genes other than *APOE* exert independent effects on AD risk. In other words, the high heritability of AD [34] may be explained by many and potentially complex interactions among many genes. However, experience from GWAS for other diseases suggests that failure to obtain robust results may be a matter of sample size. In other diseases in which GWAS have been successful (eg, type 2 diabetes, breast cancer), sample sizes between 5,000 and 10,000 cases were required to detect risk variants with small effects on disease risk (ie, ORs≤ 1.2). By comparison, the largest of the AD GWAS described above [21] had fewer than 3000 cases. Lack of consistent results across studies also may be attributed to variable definitions of cases and controls, subject ascertainment (eg, community-based samples are typically older than clinic-based populations; case status in autopsy series is determined by neuropathological findings compared with samples of living individuals, in whom case status is based on cognitive assessment and brain imaging), and SNP genotyping platforms and microarray chips.

Consortium Efforts

Recently, several consortia have been formed to combine AD GWAS datasets to increase power for detection of association and to standardize phenotype definitions, quality control procedures for genotype data, and statistical analysis methods (Table 2). The relative enormity of these study populations has enabled the creation of optimally sized discovery and replication datasets. However, the aggregation of studies with different recruitment protocols, genetic backgrounds, genotyping platforms, and SNP chips increases the possibility of false-positive and false-negative results. The common approach to minimize these errors is to analyze each dataset independently and combine the results across datasets using meta-analysis procedures that can account for differences in sample size and strength, and direction of the SNP effects on disease risk.

In 2009, two groups simultaneously published the results from their large consortia efforts [35•, 36•]. The Genetic and Environmental Risk in Alzheimer Disease (GERAD) Consortium established a discovery dataset including 3,941 cases and 7,848 controls ascertained at multiple locations in the United Kingdom, and a replication dataset including 2,023 cases and 2,340 controls obtained from cohorts in Germany, Belgium, Greece, and the United States [35•]. The European Alzheimer Disease Initiative (EADI) Consortium grouped AD GWAS cohorts from several countries in Europe into a discovery dataset including 2,032 cases and 5,328 controls from France, and a replication dataset including 3,978 cases and 3,297 controls from Italy, Spain, Belgium, and Finland [36•]. Both studies identified genome-wide significant association with SNPs in clusterin (*CLU*), also known as *APOJ*. Genome-wide significant associations were also reported for phosphatidylinositol-binding clathrin assembly protein (*PICALM*) by the GERAD Consortium [35•] and complement component (3b/4b) receptor 1 (*CR1*) by the EADI Consortium [36•]. Lambert et al. [36•] estimated the population-attributable risks for three of the primary AD genes to be 25.5% (*APOE*), 8.9% (*CLU*), and 3.8% (*CR1*). However, these estimates are inflated because they were calculated from cohorts likely to show much stronger effects for these loci

Table 2 Summary of consortium studies to identify Alzheimer's disease risk variants

Study	Recruitment sites	SNPs tested, n	Cases, n	Controls, n	Non-*APOE* genes identified	P value
Harold et al. [35•]	United States and Europe	610,000	5,964	10,188	*CLU*	1.4×10^{-9}
					PICALM	1.9×10^{-8}
Lambert et al. [36•]	Europe	537,029	6,010	8,625	*CLU*	7.5×10^{-9}
					CR1	3.7×10^{-9}
Seshadri et al. [37]	United States and Europe	2,500,000	9,738	26,965	*CLU*	1.6×10^{-16}
					PICALM	3.2×10^{-12}
					BIN1	1.6×10^{-11}
					EXOC3L2	6.4×10^{-9}
Jun et al. [38]	United States, Canada, Dominican Republic, and Israel[a]	17	7,070	8,169	*CR1*	6.1×10^{-5}
					CLU	0.0002
					PICALM	2.3×10^{-5}

[a] Results reported for non-Hispanic Caucasians of European descent only

SNP single nucleotide polymorphism

(ie, the "winner's curse," the tendency for the significance and effect size of results in a moderately powered discovery sample to be inflated simply because they were significant) that are not representative of the general population.

A total of 394 cases and 12,850 controls from 4 of the 6 large, prospective, community-based cohort studies participating in the Cohorts for Heart and Aging Research in Genetic Epidemiology (CHARGE) Consortium (http://web.chargeconsortium.com) formed the discovery stage of a third large GWAS for AD [37]. The discovery sample also included datasets from two of the previously published GWAS [22, 25]. Subsets of the GERAD and EADI datasets were used for replication. This GWAS identified highly significant associations with two new loci. Two SNPs (rs7561528 and rs744373) in one of these loci, the bridging integrator (BIN1) gene, were among the top subgenome-wide significant findings in the previous GERAD study [35•]. In the combined datasets, the evidence for association with rs744373 surpassed genome-wide significance. These investigators also reported genome-wide significant association with SNP rs597668 in the APOE region near BLOC1S3. This association remained significant after adjusting for APOE ε4 status, suggesting that the effect of this gene is independent of APOE.

The Alzheimer's Disease Genetics Consortium (ADGC) (http://alois.med.upenn.edu/adgc) confirmed associations of AD with SNPs in CLU, PICALM, and CR1 by meta-analysis of 9 Caucasian datasets independent of the previous reports comprising 5,686 cases and 5,852 controls [38]. Further analysis of these data showed that the effect of PICALM on AD risk is greater among APOE ε4 carriers. None of the tested SNPs were significantly associated with AD in modest samples of African Americans, Caribbean Hispanics, or Israeli-Arabs.

The potential for novel discovery by the ADGC is considerable. This consortium includes nearly all existing GWAS datasets in the United States and brought together the nation's resources from the Alzheimer's Disease Centers, National Alzheimer's Coordinating Center, and National Cell Repository for Alzheimer's Disease to create a de novo GWAS dataset that currently has more than 3,200 cases and 1,250 controls. This consortium assembled many large, ethnically diverse datasets containing more than 12,000 cases and 11,000 controls that will enable investigation of the genetics of AD and related traits, including neuropathological traits, measures of cognitive function and memory, rate of disease progression, and biomarkers.

Recently, the ADGC conducted a GWAS including a discovery dataset of more than 8,300 cases and 7,350 cognitively healthy older adult controls, and a replication dataset including about 3,500 cases and 3,500 controls. Genome-wide significant associations were identified with

four novel loci. It is expected that details of these findings will be published later this year.

Despite these successes in identifying AD risk variants in multiple samples, a large proportion of the genetic variation contributing to AD remains unidentified. The population-attributable risk for APOE, CLU, PICALM, and CR1 combined was estimated to be as high as 56% [10], but the true proportion of the genetic variance accounted for by these genes is certainly much lower, as this estimate assumes non-additivity of the effects and was determined from research samples that overestimate the strength of the association (ie, the "winner's curse") and do not represent the population as a whole. Using a true population sample, Harold et al. [35•] reported a very modest increase in the predictive value of a model containing age, sex, APOE, PICALM, and CLU over one containing only age, sex, and APOE.

Elucidating Biological Mechanisms of Alzheimer's Disease Through Genome-Wide Association Studies

The inherent agnostic nature of the GWAS approach facilitates the discovery of new pathways and mechanisms for AD. The associated genes or the pathways in which they have a role may provide critical insight into the development of effective strategies for disease intervention or prevention. However, despite the statistical evidence for association, the functional variants in the novel genes have not yet been identified, and the precise roles of their encoded proteins in AD pathogenesis are poorly understood.

Even the biological effect of APOE, the most widely studied and potent genetic risk factor for AD, has not been established with certainty, although many mechanisms have been proposed [39].

CLU, like APOE, is an apolipoprotein expressed at high levels in the brain and may be involved in synaptic turnover [40] and apoptosis [41]. In rats, secreted CLU is present in amyloid plaques [42]. Calero et al. [43] suggested four potential pathways through which CLU may impact AD: neuroprotection through antiapoptotic signaling; protection against oxidative stress; inhibition of the membrane attack complex of complement proteins in response to inflammation; and binding to hydrophobic regions of partially unfolded, stressed proteins, thereby preventing their aggregation. Complement component (3b/4b) receptor 1 (CR1) may be involved in Aβ clearance via the complement system. The evidence for this comes from a murine model in which overexpression of C3 in transgenic mice resulted in lower Aβ deposition and neurodegeneration was observed in the C3 knockdown mice [44].

PICALM is believed to affect AD through intracellular trafficking of Aβ. PICALM is a cofactor in clathrin-mediated endocytosis, which, in addition to Aβ, routes

proteins, lipids, growth factors, and neurotransmitters to various regions within the cell, where they can be differentially processed, secreted, or degraded. The specific function of *BIN1* in AD pathogenesis is also unclear. One possibility for *BIN1*'s role in AD is through its binding partner, integrin α-3, which mediates neuronal adhesion and migration [45] and detachment of migrating neurons from radial glial fibers in mice [46]. Other members of the *BIN1* gene family function in neuronal membrane organization and clathrin-mediated synaptic vessel formation, which can be disrupted by Aβ [47].

Genome Mining

As large consortia of GWAS achieve the power necessary to identify most of the single variants with small effects on AD risk, alternative strategies to explain the remaining heritability of AD are being developed and tested. One hypothesis is that most of the heritability has already been explained by the net effect of the thousands of SNPs associated with AD (at $5 \times 10^{-8} < P < 0.05$). If true, sample sizes of hundreds of thousands or more would be needed to separate the true from the false associations considering the likely very small effects of these SNPs on AD risk, and the disease predictive value or therapeutic target potential gained by doing so likely would be minimal.

Researchers are considering other strategies to identify additional genes that may have important roles in AD but cannot be implicated by GWAS. For example, pathway and gene set enrichment analysis are related methods that have been used to identify risk variants in genes and gene networks in other complex diseases. These methods combine data on gene association and expression, as well as other types of evidence to search for biochemical pathways including genes that show a higher-than-expected number of nominally significant single SNP associations. A recent analysis of 2,344 AD cases and 7,076 controls showed significant enrichment of AD-associated genes in several pathways [48]. As expected, the "Alzheimer's disease" pathway showed the greatest level of enrichment, but several other pathways were significant, including "regulation of autophagy," "natural killer cell–mediated cytotoxicity," "antigen processing and presentation" and "RIG-I–like receptor signaling." A similar recent study of a large French cohort identified a significant enrichment of AD associations with SNPs in genes in pathways related to transmembrane transport of nucleoporins and mitochondrial proteins [49].

It is generally assumed that interaction among genes (termed *epistasis*) within or across biological pathways contributes to AD pathology. Recently, Jiang et al. [50], using a Bayesian networks approach, confirmed a previously reported association of AD with *GAB2* that was specific to *APOE* ε4 carriers [28], albeit in the same dataset using a different methodology. Although it is perhaps not surprising that the same dataset yielded similar results using two different methods, this work highlights the ability of Bayesian network analysis to detect relevant gene–gene interactions in a computationally efficient manner.

Conclusions

Despite the disappointing results from initial GWAS, the search for AD risk loci using this approach has yielded several exciting findings. Large datasets assembled by several consortia are now providing the statistical power necessary to detect small effect loci, and the recently identified genetic variants increase risk very modestly. As the limit to the number of GWAS-detectable variants is reached, much of the heritability of AD remains unexplained. Several alternate strategies are being pursued, including next-generation sequencing to identify rare variants not captured by current genotyping platforms, mining existing SNP data for higher order interactions, and studying unique populations. These new genetic discoveries offer new clues and affirm several previous ideas about disease mechanisms, including Aβ trafficking and deposition, tangle formation, mitochondrial function, oxidative stress, lipid homeostasis dysregulation, loss of synaptic plasticity, and cholinergic and immune dysfunction. In addition, these genes are potential targets for new drugs and other interventions.

References

Papers of particular interest, published recently, have been highlighted as:
• Of importance

1. Levy-Lahad E, Wijsman EM, Nemens E, Anderson L, Goddard KA, Weber JL, et al. A familial Alzheimer's disease locus on chromosome 1. Science. 1995;269:970–3.
2. St George-Hyslop P, Haines J, Rogaev E, Mortilla M, Vaula G, Pericak-Vance M, et al. Genetic evidence for a novel familial Alzheimer's disease locus on chromosome 14. Nat Genet. 1992;2:330–4.
3. Schellenberg GD, Bird TD, Wijsman EM, Orr HT, Anderson L, Nemens E, et al. Genetic linkage evidence for a familial Alzheimer's disease locus on chromosome 14. Science. 1992;258:668–71.

4. Tanzi RE, Gusella JF, Watkins PC, Bruns GA, St George-Hyslop P, Van Keuren ML, et al. Amyloid beta protein gene: cDNA, mRNA distribution, and genetic linkage near the Alzheimer locus. Science. 1987;235:880–4.

5. Rogaev EI, Sherrington R, Rogaeva EA, Levesque G, Ikeda M, Liang Y, et al. Familial Alzheimer's disease in kindreds with missense mutations in a gene on chromosome 1 related to the Alzheimer's disease type 3 gene. Nature. 1995;376:775–8.

6. Sherrington R, Rogaev EI, Liang Y, Rogaeva EA, Levesque G, Ikeda M, et al. Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's disease. Nature. 1995;375:754–60.

7. Goate A, Chartier-Harlin MC, Mullan M, Brown J, Crawford F, Fidani L, et al. Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease. Nature. 1991;349:704–6.

8. Farrer LA, Cupples LA, Haines JL, Hyman B, Kukull WA, Mayeux R, et al. Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease. A meta-analysis. APOE and Alzheimer Disease Meta Analysis Consortium. JAMA. 1997;278:1349–56.

9. Kennedy JL, Farrer LA, Andreasen NC, Mayeux R, St George-Hyslop P. The genetics of adult-onset neuropsychiatric disease: complexities and conundra? Science. 2003;302:822–6.

10. Ertekin-Taner N. Genetics of Alzheimer disease in the pre- and post-GWAS era. Alzheimers Res Ther. 2010;2:3.

11. Pericak-Vance MA, Bebout JL, Gaskell Jr PC, Yamaoka LH, Hung WY, Alberts MJ, et al. Linkage studies in familial Alzheimer disease: evidence for chromosome 19 linkage. Am J Hum Genet. 1991;48:1034–50.

12. 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.

13. Rogaeva E, Meng Y, Lee JH, Gu Y, Kawarai T, Zou F, et al. The neuronal sortilin-related receptor SORL1 is genetically associated with Alzheimer disease. Nat Genet. 2007;39:168–77.

14. Elston RC. Linkage and association. Genet Epidemiol. 1998;15:565–76.

15. Lehmann DJ, Cortina-Borja M, Warden DR, Smith AD, Sleegers K, Prince JA, et al. Large meta-analysis establishes the ACE insertion-deletion polymorphism as a marker of Alzheimer's disease. Am J Epidemiol. 2005;162:305–17.

16. Meng Y, Baldwin CT, Bowirrat A, Waraska K, Inzelberg R, Friedland RP, et al. Association of polymorphisms in the Angiotensin-converting enzyme gene with Alzheimer disease in an Israeli Arab community. Am J Hum Genet. 2006;78:871–7.

17. Uda M, Galanello R, Sanna S, Lettre G, Sankaran VG, Chen W, et al. Genome-wide association study shows BCL11A associated with persistent fetal hemoglobin and amelioration of the phenotype of beta-thalassemia. Proc Natl Acad Sci U S A. 2008;105:1620–5.

18. Bonnefond A, Froguel P, Vaxillaire M. The emerging genetics of type 2 diabetes. Trends Mol Med. 2010;16:407–16.

19. Ghoussaini M, Song H, Koessler T, Al Olama AA, Kote-Jarai Z, Driver KE, et al. Multiple loci with different cancer specificities within the 8q24 gene desert. J Natl Cancer Inst. 2008;100:962–6.

20. Grupe A, Abraham R, Li Y, Rowland C, Hollingworth P, Morgan A, et al. Evidence for novel susceptibility genes for late-onset Alzheimer's disease from a genome-wide association study of putative functional variants. Hum Mol Genet. 2007;16:865–73.

21. Bertram L, Lange C, Mullin K, Parkinson M, Hsiao M, Hogan MF, et al. Genome-wide association analysis reveals putative Alzheimer's disease susceptibility loci in addition to APOE. Am J Hum Genet. 2008;83:623–32.

22. Coon KD, Myers AJ, Craig DW, Webster JA, Pearson JV, Lince DH, et al. A high-density whole-genome association study reveals that APOE is the major susceptibility gene for sporadic late-onset Alzheimer's disease. J Clin Psychiatry. 2007;68:613–8.

23. Li H, Wetten S, Li L, St Jean PL, Upmanyu R, Surh L, et al. Candidate single-nucleotide polymorphisms from a genomewide association study of Alzheimer disease. Arch Neurol. 2008;65:45–53.

24. Abraham R, Moskvina V, Sims R, Hollingworth P, Morgan A, Georgieva L, et al. A genome-wide association study for late-onset Alzheimer's disease using DNA pooling. BMC Med Genomics. 2008;1:44.

25. Carrasquillo MM, Zou F, Pankratz VS, Wilcox SL, Ma L, Walker LP, et al. Genetic variation in PCDH11X is associated with susceptibility to late-onset Alzheimer's disease. Nat Genet. 2009;41:192–8.

26. Beecham GW, Martin ER, Li YJ, Slifer MA, Gilbert JR, Haines JL, et al. Genome-wide association study implicates a chromosome 12 risk locus for late-onset Alzheimer disease. Am J Hum Genet. 2009;84:35–43.

27. Poduslo SE, Huang R, Huang J, Smith S. Genome screen of late-onset Alzheimer's extended pedigrees identifies TRPC4AP by haplotype analysis. Am J Med Genet B Neuropsychiatr Genet. 2009;150B:50–5.

28. Reiman EM, Webster JA, Myers AJ, Hardy J, Dunckley T, Zismann VL, et al. GAB2 alleles modify Alzheimer's risk in APOE epsilon4 carriers. Neuron. 2007;54:713–20.

29. Sherva R, Baldwin CT, Inzelberg R, Vardarajan B, Cupples LA, Lunetta K, et al. Identification of novel candidate genes for Alzheimer's disease by autozygosity mapping using genome wide SNP data from an Israeli-Arab community. J Alzheimers Dis (2010), (Epub ahead of print) PMID: 21098978.

30. Kang DE, Yoon IS, Repetto E, Busse T, Yermian N, Ie L, et al. Presenilins mediate phosphatidylinositol 3-kinase/AKT and ERK activation via select signaling receptors. Selectivity of PS2 in platelet-derived growth factor signaling. J Biol Chem. 2005;280:31537–47.

31. Liang WS, Chen K, Lee W, Sidhar K, Corneveaux JJ, Allen AN, et al. Association between GAB2 haplotype and higher glucose metabolism in Alzheimer's disease-affected brain regions in cognitively normal APOEepsilon4 carriers. Neuroimage (2010).

32. Naj AC, Beecham GW, Martin ER, Gallins PJ, Powell EH, Konidari I, et al. Dementia revealed: novel chromosome 6 locus for late-onset Alzheimer disease provides genetic evidence for folate-pathway abnormalities. PLoS Genet. 2010;6:e1001130.

33. Bowirrat A, Friedland RP, Chapman J, Korczyn AD. The very high prevalence of AD in an Arab population is not explained by APOE epsilon4 allele frequency. Neurology. 2000;55:731.

34. 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.

35. • 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. *This is one of two papers published back to back reporting results from a GWAS of AD in a very large sample composed of datasets from multiple research teams. Genome-wide significant results were obtained for variants in two novel genes whose role in the disorder would not have been predicted from their known functions. This paper demonstrates the importance of having a dataset exceeding 10,000 subjects to detect association with genes conferring a small effect on disease risk.*

36. • 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. *This is the second of two back-to-back papers reporting results from a GWAS of AD in a very large sample composed of datasets from a second consortium of AD genetics researchers. Genome-wide significant results were obtained for*

variants in two novel genes, one of which (PICALM) was also identified in an independent study.

37. 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.

38. Jun G, Naj AC, Beecham GW, Wang LS, Buros J, Gallins PJ, et al. Meta-analysis Confirms CR1, CLU, and PICALM as Alzheimer disease risk loci and reveals interactions with APOE genotypes. Arch Neurol (2010), (Epub ahead of print) PMID: 20697030.

39. Huang Y. Abeta-independent roles of apolipoprotein E4 in the pathogenesis of Alzheimer's disease. Trends Mol Med. 2010;16:287–94.

40. Danik M, Chabot JG, Hassan-Gonzalez D, Suh M, Quirion R. Localization of sulfated glycoprotein-2/clusterin mRNA in the rat brain by in situ hybridization. J Comp Neurol. 1993;334:209–27.

41. Wong P, Pineault J, Lakins J, Taillefer D, Leger J, Wang C, et al. Genomic organization and expression of the rat TRPM-2 (clusterin) gene, a gene implicated in apoptosis. J Biol Chem. 1993;268:5021–31.

42. May PC, Lampert-Etchells M, Johnson SA, Poirier J, Masters JN, Finch CE. Dynamics of gene expression for a hippocampal glycoprotein elevated in Alzheimer's disease and in response to experimental lesions in rat. Neuron. 1990;5:831–9.

43. Calero M, Rostagno A, Matsubara E, Zlokovic B, Frangione B, Ghiso J. Apolipoprotein J (clusterin) and Alzheimer's disease. Microsc Res Tech. 2000;50:305–15.

44. Wyss-Coray T, Yan F, Lin AH, Lambris JD, Alexander JJ, Quigg RJ, et al. Prominent neurodegeneration and increased plaque formation in complement-inhibited Alzheimer's mice. Proc Natl Acad Sci U S A. 2002;99:10837–42.

45. Dulabon L, Olson EC, Taglienti MG, Eisenhuth S, McGrath B, Walsh CA, et al. Reelin binds alpha3beta1 integrin and inhibits neuronal migration. Neuron. 2000;27:33–44.

46. Sanada K, Gupta A, Tsai LH. Disabled-1-regulated adhesion of migrating neurons to radial glial fiber contributes to neuronal positioning during early corticogenesis. Neuron. 2004;42:197–211.

47. Kelly BL, Ferreira A. Beta-amyloid disrupted synaptic vesicle endocytosis in cultured hippocampal neurons. Neuroscience. 2007;147:60–70.

48. Lambert JC, Grenier-Boley B, Chouraki V, Heath S, Zelenika D, Fievet N, et al. Implication of the immune system in Alzheimer's disease: evidence from genome-wide pathway analysis. J Alzheimers Dis. 2010;20:1107–18.

49. Hong MG, Alexeyenko A, Lambert JC, Amouyel P, Prince JA. Genome-wide pathway analysis implicates intracellular transmembrane protein transport in Alzheimer disease. J Hum Genet (2010).

50. Jiang X, Barmada MM, Visweswaran S. Identifying genetic interactions in genome-wide data using Bayesian networks. Genet Epidemiol. 2010;34:575–81.