Published in final edited form as:
Am J Med Genet B Neuropsychiatr Genet. 2010 July ; 153B(5): 1060–1069. doi:10.1002/ajmg.b.31078.

Genetic Pathway-Based Hierarchical Clustering Analysis of Older Adults with Cognitive Complaints and Amnestic Mild Cognitive Impairment Using Clinical and Neuroimaging Phenotypes

Chantel D. Sloan[1], **Li Shen, PhD**[2,3], **John D. West, MSEng**[2], **Heather A. Wishart, PhD**[4], **Laura A. Flashman, PhD**[4], **Laura A. Rabin, PhD**[4], **Robert B. Santulli, MD**[4], **Stephen J. Guerin, MD**[4], **C. Harker Rhodes, MD, PhD**[5], **Gregory J. Tsongalis, PhD**[5], **Thomas W. McAllister, MD**[4], **Tim A. Ahles, PhD**[6], **Stephen L. Lee, MD, PhD**[7], **Jason H. Moore, PhD**[1,8,9,10], and **Andrew J. Saykin, PsyD**[2,4,11,*]

[1]Computational Genetics Laboratory, Departments of Genetics and Community and Family Medicine, Dartmouth Medical School, Lebanon, NH

[2]Center for Neuroimaging, Department of Radiology and Imaging Sciences, Indiana University School of Medicine, Indianapolis, IN

[3]Center for Computational Biology and Bioinformatics, Indiana University School of Medicine, Indianapolis, IN

[4]Brain Imaging Laboratory, Department of Psychiatry, Dartmouth Medical School, Lebanon, NH

[5]Department of Pathology and Laboratory Medicine, Dartmouth Medical School, Lebanon, NH

[6]Department of Psychiatry and Behavioral Sciences, Memorial Sloan-Kettering Cancer Center, New York, NY

[7]Department of Medicine (Neurology), Dartmouth Medical School, Lebanon, NH

[8]Department of Computer Science, University of New Hampshire, Durham, NH

[9]Department of Computer Science, University of Vermont, Burlington, VT

[10]Translational Genomics Research Institute, Phoenix, AZ

[11]Departments of Medical and Molecular Genetics, Neurology and Psychiatry, Indiana University School of Medicine, Indianapolis, IN

Abstract

Hierarchical clustering is frequently used for grouping results in expression or haplotype analyses. These methods can elucidate patterns between measures that can then be applied to discerning their validity in discriminating between experimental conditions. Here a hierarchical clustering method is used to analyze the results of an imaging genetics study using multiple brain morphology and cognitive testing endpoints for older adults with amnestic mild cognitive impairment (MCI) or cognitive complaints (CC) compared to healthy controls (HC). The single nucleotide polymorphisms (SNPs) are a subset of those included on a larger array that are found in a reported Alzheimer's disease (AD) and neurodegeneration pathway. The results indicate that genetic models within the endpoints cluster together, while there are 4 distinct sets of SNPs that

[*]Address for Correspondence: Dr. Andrew J. Saykin, Center for Neuroimaging, Department of Radiology and Imaging Sciences, Indiana University School of Medicine, 950 West Walnut St., R2, E124, Indianapolis, IN 46202, Phone: 317-278-6947, Fax: 317-274-1067, asaykin@iupui.edu.

differentiate between the endpoints, with most significant results associated with morphology endpoints rather than cognitive testing of patients' reported symptoms. The genes found in at least one cluster are *ABCB1, APBA1, BACE1, BACE2, BCL2, BCL2L1, CASP7, CHAT, CST3, DRD3, DRD5, IL6, LRP1, NAT1,* and *PSEN2*. The greater associations with morphology endpoints suggests that changes in brain structure can be influenced by an individual's genetic background in the absence of dementia and in some cases (Cognitive Complaints group) even without those effects necessarily being detectable on commonly used clinical tests of cognition. The results are consistent with polygenic influences on early neurodegenerative changes and demonstrate the effectiveness of hierarchical clustering in identifying genetic associations among multiple related phenotypic endpoints.

Introduction

With the rapid increase in genotyping technologies and increased ability to measure multiple endophenotypes, there is often a wealth of data and statistical results to be sorted through by investigators during the course of an analysis. As a result of this "data explosion", investigators are using a variety of methods to focus analyses and facilitate the interpretation of results and design studies to confirm findings. These methods include pathway analysis, which divides a genetic dataset into groups of genes known to have some kind of biological relationship with one another. Pathway-based analysis has been discussed as a way to study interactions more closely and increase the knowledge gained from larger genome-wide studies [Rao, 2008; Wilke, Mareedu and Moore, 2008]. Reducing analysis to individual pathways is particularly helpful when multiple endpoints and genetic models are being tested. These kinds of investigations produce an array of statistical results, but there is little literature discussing the validity of further organizing or clustering those results to discern more about the biological basis of obtained associations.

The field of imaging genetics is an excellent example of an area that produces a tremendous amount of data that subsequently needs to be evaluated and interpreted. Multiple measures of brain morphology as well as tests of cognition are administered to each patient from whom genetic data is collected. The imaging genetics study analyzed here investigated amnestic MCI, a condition characterized as a transitional state between the normal cognitive effects of aging and dementia associated with Alzheimer's disease (AD) pathology [Michon, 2009; Gauthier et al., 2006; Petersen et al., 2006; Kelley and Petersen, 2007; Fleisher et al., 2007]. Patients meeting criteria for amnestic MCI have approximately 50% risk of progression to a clinical diagnosis of AD over 5 years reflecting the strong relationship between these conditions. In individuals over the age of 65, the prevalence of MCI is somewhere between 3 and 19%. Part of the reason for the discrepancy is due to the fact that MCI is heterogeneous, having multiple subtypes of varying severity as well as the application of different diagnostic criteria [Gauthier et al., 2006]. Neuroimaging is an important biomarker with regard to risk of progression in that patients with amnestic MCI who also have reduced hippocampal volume are more likely to convert from MCI to dementia e.g.,[Jack et al., 1999; Risacher et al., 2009]. MCI patients perform differently from early probable AD patients on memory and especially on neuropsychological tests of other cognitive domains and they also may show different patterns of brain activation on functional MRI, but a clear clinical distinction between MCI and early AD remains a source of controversy [Machulda et al., 2003]. Older individuals with amnestic MCI and those with cognitive complaints (CC) in the absence of significant psychometric deficits report similar declines in cognitive function, however, those classified as CC do not have the requisite performance deficits that would lead to a diagnosis of MCI [Saykin et al., 2006]. There is evidence that CC and amnestic MCI (the subtype most likely to proceed to clinical AD) show similar patterns of brain atrophy, particularly loss of grey matter (GM) density in the

hippocampal region [Saykin et al., 2006]. Preliminary results from our group and others suggest individuals with cognitive complaints have an elevated rate of progression to MCI or dementia compared to healthy controls.

There are likely to be genetic polymorphisms that predispose individuals to develop MCI and/or CC due to the association of these conditions with AD which has known strong genetic contributors. Despite the many genetic studies of AD, there has been little research directed toward determining what specific polymorphisms are associated with MCI or CC. However, it is known that amnestic-type MCI (where a memory deficit is the predominant cognitive feature), like AD, is associated with the *APOE* ε4 allele, commonly considered the main genetic contributor to the development of late onset AD [Lopez et al., 2003; Bertram and Tanzi, 2008; Bertram et al., 2007; Farrer et al., 1997]. Cystatin C (*CST3*) and cholesterol 24-hydroxylase (*CYP46A1*) genes have been suggested to be associated with MCI among those with white matter lesions [Galluzzi et al., 2008]. A common polymorphism in the *KIBRA* gene that is associated with poorer episodic memory was not found to be associated with MCI risk [Almeida et al., 2008]. Though endophenotypes such as serum BDNF levels have been associated with MCI, individual polymorphisms such as *BDNF Val66Met* are not necessarily also associated due to the complexity of genetic interactions [Yu et al., 2008]. MCI and CC, like AD, are likely influenced by multiple interacting genes. Therefore, individual polymorphisms do not supply as much knowledge concerning the disease as pathways that incorporate ensembles of related genes and proteins.

The value of hierarchical clustering has been well established as part of various genetic and expression studies as a way to organize genes into groups that function in similar ways under certain conditions [Eisen et al., 1998; Levenstien, Yang and Ott, 2003; Sorlie et al., 2003]. For the purposes of this study we selected single nucleotide polymorphisms (SNPs) from pathways hypothesized to be involved in neurodegenerative disorders. These SNPs were then analyzed for association with neuroimaging phenotypes of CC and MCI individuals using a variety of genetic models. SNPs with at least one significant result below 0.01 were then hierarchically clustered to determine what associations were frequently associated with the same phenotypes. Therefore, we were able to investigate what groups of genetic models, rather than individual polymorphisms were associated with each of the endpoints.

Methods

Methods for recruitment, assessment and neuroimaging have been previously described in detail [Saykin et al., 2006]. A brief overview is provided here:

Recruitment of cases and controls and assessment of disease status

Subjects (n = 132) were enrolled individuals classified at baseline in the Dartmouth Memory and Aging Study as having cognitive complaints (CC) or amnestic MCI based on the criteria specified below, or as a healthy control (HC). For this analysis, subjects were considered cases if they were determined to be MCI or CC patients (n=85), while there were 47 HC participants. Each subject had to be at least 60 years of age, right handed, fluent in English, and have at least 12 years of formal education or a GED. An informant was available for each participant who knew the participant well and could answer questions about his or her cognitive and functional status. Those with any significant neurological, medical (including head trauma) or psychiatric condition other than those being investigated here were not included, as well as those with a history of substance dependence or who demonstrated nonamnestic MCI. All participants provided written informed consent according to procedures approved by the Institutional Committee for the Protection of Human Subjects.

Basic characteristics of the case and control samples with regard to covariates as well as morphological and cognitive measures are shown in Table 1.

Each participant received a detailed neuropsychological evaluation, including measures of memory, attention, executive function, language, spatial ability, general intellectual ability, and psychomotor speed as well as standard dementia screens. A geropsychiatrist (R.B.S) used a semi-structured evaluation to rule out depression, dementia and other Diagnostic and Statistical Manual of Mental Disorders, 4th Edition (DSM-IV) Axis I psychiatric disorders. A neurologic examination was also completed. Structural brain MRI scans, described below, were reviewed by a neuroradiologist blinded to clinical status, to rule out incidental pathology.

The results of the neuropsychological assessment, self and informant report indices, and the geropsychiatric and neurologic evaluation were used to classify individuals as MCI, CC or HC. The decision to characterize a participant as MCI, CC, HC or Other, was determined by a multidisciplinary consensus conference considering all available clinical information excluding quantitative MRI and genetic results. Typically, those considered to have significant CCs endorsed 20% or more of the items on a Cognitive Complaint Index (CCI) [Saykin et al., 2006]. The CC and MCI groups did not differ from each other on the CCI and endorsed approximately three times as many complaints as the HC group. Assessment of memory performance was based on age, education, and gender-adjusted scores using the mean, SD, and beta coefficients obtained from an expanded healthy demographically balanced control group. The MCI participants performed 1.5 SDs below the adjusted mean of HCs on at least one verbal memory test score (CVLT Total 1–5, Short Delay, Long Delay, WMS-III LM I or LM-II). The mean Z scores of each of these five measures were used to calculate a composite verbal memory *Z* score.

Measurement of Endpoints-clinical, neuroimaging

Scans were obtained on a GE Signa 1.5-T Horizon LX magnet, and hippocampal volume was obtained as previously described [Saykin et al., 2006; McHugh et al., 2007]. Briefly, a T1-weighted three-dimensional spoiled gradient echo (SPGR) coronal volume was acquired. Left and right hippocampal volumes were calculated for each participant by summing the coronal slice areas and then adjusted for age and total intracranial volume using a regression model. GM density within the hippocampus was measured using a region of interest (ROI) template to extract the mean value from voxel-based morphometry derived GM images. [Saykin et al., 2006].

Genotyping

Genotyping was done using a 3300 SNP targeted neuroscience panel developed in partnership between the Dartmouth Neurogenetics Group (DNG) and Affymetrix (Affymetrix Inc., Santa Clara, CA). Candidate genes and alleles were included after extensive literature review, database searching and deliberation by the DNG. Genes or alleles were included if they had a known or hypothesized involvement in critical pathways relevant to cognition or major neurological and psychiatric disorders (e.g. neurodegeneration, inflammation, depression). Candidate genes or alleles were assigned relative importance based on their probable role in neuropsychiatric processes, a reported minor allele frequency of at least 15%, and known or suspected roles in gene expression. The resulting array was implemented by Affymetrix using molecular inversion probe technology and consisted of 3300 single nucleotide polymorphisms (SNPs) from approximately 1000 candidate genes. AD/MCI is one of several disorders being investigated using this microarray. Approximately 1000 individuals have been successfully genotyped to date. Reliability was excellent, with an overall genotyping rate of 99.79%.

A pathway-based analysis was undertaken using suspected AD and neurodegeneration susceptibility genes. The SNPs selected were those from genes in the KEGG Alzheimer's disease pathway, Genetic Association Database (GAD) and AlzGene database that were available on the array (n = 110 genes, 834 SNPs)[Kanehisa et al., 2008; Becker et al., 2004; Bertram et al., 2007](Table S1). These were selected to assemble one larger Alzheimer's and neurodegeneration pathway. Many of the SNPs labeled as influencing neurodegeneration in GAD and AlzGene were also found in the KEGG Alzheimer's pathway. The creation of the larger pathway was to supplement the SNP list with genes that have strongly suggested roles in AD or MCI, but that are not yet included in KEGG.

Analysis

Before proceeding with further analyses, F-statistics were calculated using *Powermarker* to determine if genetic structure was present in the data that might bias the results [Liu and Muse, 2005]. Once the possibility of structure bias was eliminated, cases defined as those with CC or MCI were compared to healthy controls. Linear regression was used to determine the relationship between 5 different genetic models (Main Effect, Dominant, Recessive, Additive, Heterosis) per SNP and each endpoint (verbal memory score, right and left hippocampal volume and gray matter density, and the CCI). Linear regression was performed using R (v.2.7.2) [R Development Core Team, 2005] for each SNP and each model resulting in a total of 4170 statistical tests. *APOE* epsilon 4 status (ε4+ vs ε 4−) was used as a covariate for all analyses due to its already known very strong effect. For each potential endpoint, some normalizations were performed during data collection or preprocessing. Relevant factors that were not already adjusted for were included as covariates in the analysis as follows: no additional covariates were taken into account for verbal memory score, right and left hippocampal volume were covaried for age and total intracranial volume (ICV), gray matter density was covaried for age and sex, and the CCI was covaried for age, education and sex.

The linear regression results were filtered to keep only those SNPs that had at least one significant association with an endpoint and model at an α of 0.01 (given multiple testing 42 models are expected by chance). For this stage of the analysis it was determined that the α used as a threshold to filter the results should be between 0.05 and 0.01 in order to minimize false discoveries without eliminating the majority of the results. Also given the exploratory nature of this analysis with a focus on preventing type II error than type I error, 0.01 was determined to be acceptable. The results were then clustered using the R package pvclust by both SNPs and endpoints. First, hierarchical clustering is performed on each row and column using the the R package hclust, wherein the rows and columns are grouped based on their similarity. Subsequently, pvclust estimates a p-value for the clusters using bootstrap resampling [Suzuki and Shimodaira, 2006]. We chose Manhattan distance and an average agglomeration method (hclust) with 500 bootstrap replications. An α of 0.05 was selected (with rounding) to determine significance. It is important to note that the clustering, though only including SNPs with a significant association with an endpoint, was done across all the association results for each of those SNPs for each model. To further determine if the clusters of SNPs independently predict the same endpoints or if the clusters were due to linkage disequilibrium (LD), pairwise linkage disequilibrium (D') was calculated using *Powermarker* [Liu and Muse, 2005].

The genes selected for analysis had previously demonstrated some relation to AD or neurodegeneration, which increases the probability that they interact with one another. In order to further investigate if similar association patterns between genes were due to direct interactions between genes, the genes in statistically significant clusters that were not explained by LD were analyzed using Chilibot (a Natural Language Processing tool that searches for co-occurrences in Medline) [Chen and Sharp, 2004].

Results

F-statistics calculated across all of the SNPs showed only minute structure, which was not unbalanced between cases and controls ($\theta = 0.0024$); therefore further analyses did not require correction for structure bias. Within the AD and neurodegeneration pathways defined above, there were 78 SNPs in 33 genes that were significant for at least one genetic model for one endpoint at an α of 0.01, exceeding the results expected by chance. Every genetic model had at least one SNP that was significant. Figure 1a demonstrates the results of the clustering. A pattern was apparent wherein genetic models within endpoints cluster together, except for right and left hippocampal volume, for which all the models cluster together (significant clustering shown as red boxes). Figure S1 shows the dendrogram for the clustering of the endpoints. The most frequent significance was found within the models for left and right hippocampal volume and for gray matter density. There was significant clustering among SNPs as well. Several pairs or triplets of SNPs within genes significantly cluster together, though several of these can be explained by LD, as shown in the adjacent plot. There were, however, 4 larger groups (Figure 1b) of SNPs that clustered together that can only be partially explained by LD. The first (cluster 1) contained the genes *BACE1, BACE2, APBA1,* and *CST3,* which predominantly showed association with gray matter density. There was very little LD between genes in this group, though there was considerable LD within the *APBA1* gene. The second cluster had very little LD, and consisted of the SNPS within *IL6, CASP7, CHAT, APBA1,* and *LRP*. Cluster 2 appeared to be most closely associated with gray matter density and hippocampal volume. Cluster 3 contained four genes, namely *PSEN2, BACE2, ABCB1, NAT1* that were again significant in models of gray matter density and recessive and heterosis models of hippocampal volume. Cluster 4 was partially explained by LD between SNPs in the *BCL2L1* gene, but also contained a polymorphism in *BACE1, BCL2* and *DRD3* and associations with left and right hippocampal volume.

A literature search confirmed that many of the genes found in each of the clusters have a variety of roles in the cleavage of amyloid proteins, apoptosis, and inflammation (Table 2). The functions of proteins were not specific to clusters (i.e. apoptosis and amyloid related proteins were found in several clusters). Text Mining with Chilibot was therefore used to investigate interactions between genes (Table 3). Cluster 1 contained the strongest known interactions, though cluster 2 showed some evidence of possible interactions as well. Clusters 3 and 4 showed little evidence of known biological interaction. Especially strong within cluster 1 were interactions between *BACE1* and *BACE2, BACE1* and *APBA1, APBA1* and *CST3,* and *APBA1* and *BCL2*. Within cluster 2 the most evidence for biological interaction was found between *IL6* and *CHAT, APBA1* and *LRP1* and *CHAT* and *APBA1*.

Discussion

MRI has recently been confirmed as a useful measure in genetic studies of AD [Jack et al., 1999; Potkin et al., 2009; Grundman et al., 2003; Du et al., 2006] and particularly in MCI as a predictor of progression to AD with greater consistency than cognitive testing [Jack et al., 2003]. The methodology employed here, use of hierarchical clustering within genetic models and a variety of cognitive and neuroimaging endpoints, revealed many interesting potential associations. The fact that there were 78 SNPs all with a significant result <0.01 suggests that there are benefits to studying associations on a pathway level in order to more fully explore the biology behind the results of the strongest candidate genes, though a few of these are likely false positives.

The selection of raw p-values as a clustering endpoint has advantages and disadvantages. P-values were a first choice for the introduction of this method as they are readily used by

most investigators as measures of the importance of model variables. However, other metrics could likewise be explored and the clustering improved. For example, it is possible that clustering regression coefficients in fewer dimensions may introduce even more stability, while various transformations could also be applied to the data to reduce noise.

There did not appear to be LD blocks that were associated with large variations in phenotypes or inconsistent within-block patterns of association, which further serves to strengthen the validity of the clustering metric. When undertaking pathway-based analysis, the selection of the pathway is an essential step. The use of SNPs selected on the basis of several highly overlapping AD and neurodegeneration pathways appeared especially informative with regard to brain morphology as these pathways would be expected to directly influence morphological changes.

Further investigation of the pathway results using hierarchical clustering revealed which clinical measures and imaging endophenotypes were most closely associated with the measured SNPs. In this study, SNPs were frequently associated with right and left hippocampal volume and gray matter density, and rarely associated with verbal memory score or the cognitive complaint index. This suggests that changes in brain morphology are a good endophenotype for SNPs commonly associated with AD and neurodegeneration, but that these effects are not necessarily as strongly associated with cognitive tests and symptom measures in MCI and CC patients, perhaps because the morphological changes can be directly impacted by the polymorphisms effect on gene expression and protein function. This methodology further revealed directions for investigation into interactions between genes that cluster together. The genes within the clusters may impact the endpoints through different mechanisms, or there may be true additive effects or interactions that have yet to be elucidated.

Most of the genes found within clusters are associated with either beta-amyloid production or apoptosis. The remaining genes belong to signaling, immunological or metabolic pathways. The influence of beta-amyloid production on apoptosis and therefore neurodegeneration in the hippocampus has been well established [Watt et al., 1994], though the precise molecular mechanism by which beta-amyloid toxicity acts is not fully known. There are, however, several current theories centered around different protein malfunctions, including involvement of mitochondrial regulation of calcium homeostasis and interactions between caspases [Gorman, 2008; Calissano, Matrone and Amadoro, 2009; Nilsen et al., 2006]. The particular frequency with which genes involved in beta-amyloid production and apoptosis cluster with hippocampal volume and GM metrics may be due to their direct biological link to atrophy in these regions of the brain.

Several of the genes that grouped together showed considerable evidence for being biologically associated with one another. The genes in cluster 1 showed evidence for biological interaction and lend support to the validity of the hierarchical clustering approach. Homologs, such as *BACE1* and *BACE2* in cluster 1 would be expected to similarly affect morphology, as would genes that directly bind such as *BACE2* and *BCL2* and *BACE2* and *APBA1*. Within cluster 2 there was also evidence for direct biological interactions between *IL6* and *CHAT*, *APBA1* and *CHAT* and *APBA1* and *LRP1*. The two smaller clusters (3 and 4) showed little or no evidence of biological interaction, while the larger clusters seem to be associated due to connections with a few "key" genes. These genes for cluster 1 are *BACE1* and *APBA1*, and *CHAT* and *APBA1* for cluster 2, indicating that cluster 1 is largely centered around amyloid-beta production and deposition and apoptosis. Cluster 2 is centered on amyloid-beta and acetylcholine production.

As previously described, there has been much debate on the benefits and drawbacks of whole-genome vs. pathway approaches to discerning genetic effects. Though each method has advantages and disadvantages, our results suggest that pathway approaches have the potential to elucidate genetic effects using methods that may be too computationally intensive to conduct on a genome-wide scale and for which results would be difficult to interpret. Additionally, the burden of applying a threshold needed for assessing genome wide significance in the context of modest sample sizes typical for imaging genetics experiments would likely eliminate potentially important signal from candidate loci. Using hierarchical clustering to investigate similarity in results between multiple endpoints reveals added information with regard to the usefulness of those endpoints in relation to genetic background. In the present study, nearly twice the number of significant results was obtained than would be expected by chance and hierarchical clustering assisted in the organization of findings and assessment of biological plausibility. Though these results are encouraging, further investigation and replication will be required to confirm the associations between these genomic markers and brain morphology endpoints in those at risk for AD.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported in part by grants from the National Institute on Aging (R01 AG19771; P30 AG10133-18S1), Alzheimer's Association (IIRG-99-1653 from the Hedco Foundation), Indiana Economic Development Corporation (IEDC 87884), Hitchcock Foundation, and the National Alliance for Medical Image Computing (U54 EB005149). The authors gratefully acknowledge the help of Mat Falkowski and Hannah Viernes of Affymetrix, Inc. for assistance with microarray assay design and execution, Tara McHugh for performing hippocampal segmentation, Dr. Alexander Mamourian for MRI scan readings, Dr. Tor Tosteson for statistical advice, Drs. Nadia Pare and Jessica Englert, and Heather S. Pixley, Katherine Nutter-Upham, Margaret Nordstrom and Joshua Curtain for assistance with cohort assessment and follow-up, and Dr. Alan I. Green for his support of the neurogenetics project.

References

Almeida OP, Schwab SG, Lautenschlager NT, Morar B, Greenop KR, Flicker L, Wildenauer D. KIBRA genetic polymorphism influences episodic memory in later life, but does not increase the risk of mild cognitive impairment. J Cell Mol Med. 2008; 12:1672–1676. [PubMed: 18194457]

Becker KG, Barnes KC, Bright TJ, Wang SA. The genetic association database. Nat Genet. 2004; 36:431–432. [PubMed: 15118671]

Bertram L, McQueen M, Mullin K, Blacker D, Tanzi R. The AlzGene database. alzheimer research forum.

Bertram L. Alzheimer's disease genetics current status and future perspectives. Int Rev Neurobiol. 2009; 84:167–184. [PubMed: 19501718]

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. [PubMed: 17192785]

Bertram L, Tanzi RE. Thirty years of alzheimer's disease genetics: The implications of systematic meta-analyses. Nat Rev Neurosci. 2008; 9:768–778. [PubMed: 18802446]

Burgess A, Saini S, Weng YQ, Aubert I. Stimulation of choline acetyltransferase by C3d, a neural cell adhesion molecule ligand. J Neurosci Res. 2009; 87:609–616. [PubMed: 18855942]

Calissano P, Matrone C, Amadoro G. Apoptosis and in vitro alzheimer disease neuronal models. Commun Integr Biol. 2009; 2:163–169. [PubMed: 19513272]

Chen H, Sharp BM. Content-rich biological network constructed by mining PubMed abstracts. BMC Bioinformatics. 2004; 5:147. [PubMed: 15473905]

Cosentino M, Colombo C, Mauri M, Ferrari M, Corbetta S, Marino F, Bono G, Lecchini S. Expression of apoptosis-related proteins and of mRNA for dopaminergic receptors in peripheral blood mononuclear cells from patients with alzheimer disease. Alzheimer Dis Assoc Disord. 2009; 23:88–90. [PubMed: 19266703]

Cuello AC, Bruno MA, Bell KF. NGF-cholinergic dependency in brain aging, MCI and alzheimer's disease. Curr Alzheimer Res. 2007; 4:351–358. [PubMed: 17908036]

Du AT, Schuff N, Chao LL, Kornak J, Jagust WJ, Kramer JH, Reed BR, Miller BL, Norman D, Chui HC, Weiner MW. Age effects on atrophy rates of entorhinal cortex and hippocampus. Neurobiol Aging. 2006; 27:733–740. [PubMed: 15961190]

Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A. 1998; 95:14863–14868. [PubMed: 9843981]

Farrer LA, Cupples LA, Haines JL, Hyman B, Kukull WA, Mayeux R, Myers RH, Pericak-Vance MA, Risch N, van Duijn CM. Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and alzheimer disease. A meta-analysis. APOE and alzheimer disease meta analysis consortium. JAMA. 1997; 278:1349–1356. [PubMed: 9343467]

Fleisher AS, Sowell BB, Taylor C, Gamst AC, Petersen RC, Thal LJ. Alzheimer's Disease Cooperative Study. Clinical predictors of progression to alzheimer disease in amnestic mild cognitive impairment. Neurology. 2007; 68:1588–1595. [PubMed: 17287448]

Galluzzi S, Geroldi C, Benussi L, Ghidoni R, Testa C, Borsci G, Bonetti M, Manfellotto D, Romanelli G, Zulli R, Binetti G, Frisoni GB. Association of blood pressure and genetic background with white matter lesions in patients with mild cognitive impairment. J Gerontol A Biol Sci Med Sci. 2008; 63:510–517. [PubMed: 18511756]

Gauthier S, Reisberg B, Zaudig M, Petersen RC, Ritchie K, Broich K, Belleville S, Brodaty H, Bennett D, Chertkow H, Cummings JL, de Leon M, Feldman H, Ganguli M, Hampel H, Scheltens P, Tierney MC, Whitehouse P, Winblad B. International Psychogeriatric Association Expert Conference on mild cognitive impairment. Mild cognitive impairment. Lancet. 2006a; 367:1262–1270. [PubMed: 16631882]

Gauthier S, Reisberg B, Zaudig M, Petersen RC, Ritchie K, Broich K, Belleville S, Brodaty H, Bennett D, Chertkow H, Cummings JL, de Leon M, Feldman H, Ganguli M, Hampel H, Scheltens P, Tierney MC, Whitehouse P, Winblad B. International Psychogeriatric Association Expert Conference on mild cognitive impairment. Mild cognitive impairment. Lancet. 2006b; 367:1262–1270. [PubMed: 16631882]

Gorman AM. Neuronal cell death in neurodegenerative diseases: Recurring themes around protein handling. J Cell Mol Med. 2008; 12:2263–2280. [PubMed: 18624755]

Grundman M, Jack CR Jr, Petersen RC, Kim HT, Taylor C, Datvian M, Weiner MF, DeCarli C, DeKosky ST, van Dyck C, Darvesh S, Yaffe K, Kaye J, Ferris SH, Thomas RG, Thal LJ. Alzheimer's Disease Cooperative Study. Hippocampal volume is associated with memory but not monmemory cognitive performance in patients with mild cognitive impairment. J Mol Neurosci. 2003; 20:241–248. [PubMed: 14501003]

Haass C. Take five--BACE and the gamma-secretase quartet conduct alzheimer's amyloid beta-peptide generation. EMBO J. 2004; 23:483–488. [PubMed: 14749724]

Ho A, Liu X, Sudhof TC. Deletion of mint proteins decreases amyloid production in transgenic mouse models of alzheimer's disease. J Neurosci. 2008; 28:14392–14400. [PubMed: 19118172]

Jack CR Jr, Petersen RC, Xu YC, O'Brien PC, Smith GE, Ivnik RJ, Boeve BF, Waring SC, Tangalos EG, Kokmen E. Prediction of AD with MRI-based hippocampal volume in mild cognitive impairment. Neurology. 1999; 52:1397–1403. [PubMed: 10227624]

Jack CR Jr, Slomkowski M, Gracon S, Hoover TM, Felmlee JP, Stewart K, Xu Y, Shiung M, O'Brien PC, Cha R, Knopman D, Petersen RC. MRI as a biomarker of disease progression in a therapeutic trial of milameline for AD. Neurology. 2003; 60:253–260. [PubMed: 12552040]

Johnson N, Bell P, Jonovska V, Budge M, Sim E. NAT gene polymorphisms and susceptibility to alzheimer's disease: Identification of a novel NAT1 allelic variant. BMC Med Genet. 2004; 5:6. [PubMed: 15142281]

Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, Katayama T, Kawashima S, Okuda S, Tokimatsu T, Yamanishi Y. KEGG for linking genomes to life and the environment. Nucleic Acids Res. 2008; 36:D480–D484. [PubMed: 18077471]

Kelley BJ, Petersen RC. Alzheimer's disease and mild cognitive impairment. Neurol Clin. 2007; 25:577–609. v. [PubMed: 17659182]

Kuhn PH, Marjaux E, Imhof A, De Strooper B, Haass C, Lichtenthaler SF. Regulated intramembrane proteolysis of the interleukin-1 receptor II by alpha-, beta-, and gamma-secretase. J Biol Chem. 2007; 282:11982–11995. [PubMed: 17307738]

Kuhnke D, Jedlitschky G, Grube M, Krohn M, Jucker M, Mosyagin I, Cascorbi I, Walker LC, Kroemer HK, Warzok RW, Vogelgesang S. MDR1-P-glycoprotein (ABCB1) mediates transport of alzheimer's amyloid-beta peptides--implications for the mechanisms of abeta clearance at the blood-brain barrier. Brain Pathol. 2007; 17:347–353. [PubMed: 17610523]

Lam FC, Liu R, Lu P, Shapiro AB, Renoir JM, Sharom FJ, Reiner PB. Beta-amyloid efflux mediated by p-glycoprotein. J Neurochem. 2001; 76:1121–1128. [PubMed: 11181832]

Levenstien MA, Yang Y, Ott J. Statistical significance for hierarchical clustering in genetic association and microarray expression studies. BMC Bioinformatics. 2003; 4:62. [PubMed: 14667254]

Lin C, Wang ST, Wu CW, Chuo LJ, Kuo YM. The association of a cystatin C gene polymorphism with late-onset alzheimer's disease and vascular dementia. Chin J Physiol. 2003; 46:111–115. [PubMed: 14672279]

Liu K, Muse SV. PowerMarker:Integrated analysis environment for genetic marker data. Bioinformatics. 2005:2128–2129. [PubMed: 15705655]

Lopez OL, Jagust WJ, Dulberg C, Becker JT, DeKosky ST, Fitzpatrick A, Breitner J, Lyketsos C, Jones B, Kawas C, Carlson M, Kuller LH. Risk factors for mild cognitive impairment in the cardiovascular health study cognition study: Part 2. Arch Neurol. 2003:1394–1399. [PubMed: 14568809]

Lu G, Kwong WH, Li Q, Wang X, Feng Z, Yew DT. Bcl2, bax, and nestin in the brains of patients with neurodegeneration and those of normal aging. J Mol Neurosci. 2005; 27:167–174. [PubMed: 16186627]

Lukiw WJ, Bazan NG. Survival signalling in alzheimer's disease. Biochem Soc Trans. 2006; 34:1277–1282. [PubMed: 17073801]

Machulda MM, Ward HA, Borowski B, Gunter JL, Cha RH, O'Brien PC, Petersen RC, Boeve BF, Knopman D, Tang-Wai DF, Ivnik RJ, Smith GE, Tangalos EG, Jack CR Jr. Comparison of memory fMRI response among normal, MCI, and alzheimer's patients. Neurology. 2003; 61:500–506. [PubMed: 12939424]

Marcon G, Di Fede G, Giaccone G, Rossi G, Giovagnoli AR, Maccagnano E, Tagliavini F. A novel italian presenilin 2 gene mutation with prevalent behavioral phenotype. J Alzheimers Dis. 2009; 16:509–511. [PubMed: 19276543]

Matsui T, Ramasamy K, Ingelsson M, Fukumoto H, Conrad C, Frosch MP, Irizarry MC, Yuan J, Hyman BT. Coordinated expression of caspase 8, 3 and 7 mRNA in temporal cortex of alzheimer disease: Relationship to formic acid extractable abeta42 levels. J Neuropathol Exp Neurol. 2006; 65:508–515. [PubMed: 16772874]

McConlogue L, Buttini M, Anderson JP, Brigham EF, Chen KS, Freedman SB, Games D, Johnson-Wood K, Lee M, Zeller M, Liu W, Motter R, Sinha S. Partial reduction of BACE1 has dramatic effects on alzheimer plaque and synaptic pathology in APP transgenic mice. J Biol Chem. 2007; 282:26326–26334. [PubMed: 17616527]

McHugh TL, Saykin AJ, Wishart HA, Flashman LA, Cleavinger HB, Rabin LA, Mamourian AC, Shen L. Hippocampal volume and shape analysis in an older adult population. Clin Neuropsychol. 2007; 21:130–145. [PubMed: 17366281]

Michon A. The concept of mild cognitive impairment: Relevance and limits in clinical practice. Front Neurol Neurosci. 2009; 24:12–19. [PubMed: 19182458]

Miller CC, McLoughlin DM, Lau KF, Tennant ME, Rogelj B. The X11 proteins, abeta production and alzheimer's disease. Trends Neurosci. 2006; 29:280–285. [PubMed: 16545469]

Monastero R, Camarda C, Cefalu AB, Caldarella R, Camarda LK, Noto D, Averna MR, Camarda R. No association between the cystatin C gene polymorphism and alzheimer's disease: A case-control study in an italian population. J Alzheimers Dis. 2005; 7:291–295. [PubMed: 16131730]

Nacmias B, Bagnoli S, Tedde A, Cellini E, Guarnieri BM, Bartoli A, Serio A, Piacentini S, Sorbi S. Cystatin C and apoe polymorphisms in italian alzheimer's disease. Neurosci Lett. 2006; 392:110–113. [PubMed: 16188386]

Nilsen J, Chen S, Irwin RW, Iwamoto S, Brinton RD. Estrogen protects neuronal cells from amyloid beta-induced apoptosis via regulation of mitochondrial proteins and function. BMC Neurosci. 2006; 7:74. [PubMed: 17083736]

Papassotiropoulos A, Hock C, Nitsch RM. Genetics of interleukin 6: Implications for alzheimer's disease. Neurobiol Aging. 2001; 22:863–871. [PubMed: 11754993]

Petersen RC, Parisi JE, Dickson DW, Johnson KA, Knopman DS, Boeve BF, Jicha GA, Ivnik RJ, Smith GE, Tangalos EG, Braak H, Kokmen E. Neuropathologic features of amnestic mild cognitive impairment. Arch Neurol. 2006; 63:665–672. [PubMed: 16682536]

Pompl PN, Yemul S, Xiang Z, Ho L, Haroutunian V, Purohit D, Mohs R, Pasinetti GM. Caspase gene expression in the brain as a function of the clinical progression of alzheimer disease. Arch Neurol. 2003; 60:369–376. [PubMed: 12633148]

Potkin SG, Guffanti G, Lakatos A, Turner JA, Kruggel F, Fallon JH, Saykin AJ, Orro A, Lupoli S, Salvi E, Weiner M, Macciardi F. for the Alzheimer's Disease Neuroimaging Initiative. Hippocampal atrophy as a quantitative trait in a genome-wide association study identifying novel susceptibility genes for alzheimer's disease. PLoS One. 2009; 4:e6501. [PubMed: 19668339]

Quintanilla RA, Orellana DI, Gonzalez-Billault C, Maccioni RB. Interleukin-6 induces alzheimer-type phosphorylation of tau protein by deregulating the cdk5/p35 pathway. Exp Cell Res. 2004; 295:245–257. [PubMed: 15051507]

R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria: Foundation for Statistical Computing; 2005.

Rao DC. An overview of the genetic dissection of complex traits. Adv Genet. 2008; 60:3–34. [PubMed: 18358314]

Risacher SL, Saykin AJ, West JD, Shen L, Firpi HA, McDonald BC. Alzheimer's Disease Neuroimaging Initiative (ADNI). Baseline MRI predictors of conversion from MCI to probable AD in the ADNI cohort. Current Alzheimer Research. 2009; 6:347–361. [PubMed: 19689234]

Saito Y, Sano Y, Vassar R, Gandy S, Nakaya T, Yamamoto T, Suzuki T. X11 proteins regulate the translocation of amyloid beta-protein precursor (APP) into detergent-resistant membrane and suppress the amyloidogenic cleavage of APP by beta-site-cleaving enzyme in brain. J Biol Chem. 2008; 283:35763–35771. [PubMed: 18845544]

Saykin AJ, Wishart HA, Rabin LA, Santulli RB, Flashman LA, West JD, McHugh TL, Mamourian AC. Older adults with cognitive complaints show brain atrophy similar to that of amnestic MCI. Neurology. 2006; 67:834–842. [PubMed: 16966547]

Serretti A, Olgiati P, De Ronchi D. Genetics of alzheimer's disease. A rapidly evolving field. J Alzheimers Dis. 2007; 12:73–92. [PubMed: 17851196]

Shimohama S. Nicotinic receptor-mediated neuroprotection in neurodegenerative disease models. Biol Pharm Bull. 2009; 32:332–336. [PubMed: 19252273]

Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, Deng S, Johnsen H, Pesich R, Geisler S, Demeter J, Perou CM, Lonning PE, Brown PO, Borresen-Dale AL, Botstein D. Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci U S A. 2003; 100:8418–8423. [PubMed: 12829800]

Stockley JH, O'Neill C. The proteins BACE1 and BACE2 and beta-secretase activity in normal and alzheimer's disease brain. Biochem Soc Trans. 2007; 35:574–576. [PubMed: 17511655]

Suzuki R, Shimodaira H. Pvclust: An R package for assessing the uncertainty in hierarchical clustering. Bioinformatics. 2006; 22:1540–1542. [PubMed: 16595560]

Waldron E, Heilig C, Schweitzer A, Nadella N, Jaeger S, Martin AM, Weggen S, Brix K, Pietrzik CU. LRP1 modulates APP trafficking along early compartments of the secretory pathway. Neurobiol Dis. 2008; 31:188–197. [PubMed: 18559293]

Watt JA, Pike CJ, Walencewicz-Wasserman AJ, Cotman CW. Ultrastructural analysis of beta-amyloid-induced apoptosis in cultured hippocampal neurons. Brain Res. 1994; 661:147–156. [PubMed: 7834365]

Wilke RA, Mareedu RK, Moore JH. The pathway less traveled: Moving from candidate genes to candidate pathways in the analysis of genome-wide data from large scale pharmacogenetic association studies. Curr Pharmacogenomics Person Med. 2008; 6:150–159. [PubMed: 19421424]

Willem M, Lammich S, Haass C. Function, regulation and therapeutic properties of beta-secretase (BACE1). Semin Cell Dev Biol. 2009; 20:175–182. [PubMed: 19429494]

Yamada K, Hashimoto T, Yabuki C, Nagae Y, Tachikawa M, Strickland DK, Liu Q, Bu G, Basak JM, Holtzman DM, Ohtsuki S, Terasaki T, Iwatsubo T. The low density lipoprotein receptor-related protein 1 mediates uptake of amyloid beta peptides in an in vitro model of the blood-brain barrier cells. J Biol Chem. 2008; 283:34554–34562. [PubMed: 18940800]

Yu H, Zhang Z, Shi Y, Bai F, Xie C, Qian Y, Yuan Y, Deng L. Association study of the decreased serum BDNF concentrations in amnestic mild cognitive impairment and the Val66Met polymorphism in Chinese Han. J Clin Psychiatry. 2008; 69:1104–1111. [PubMed: 18505307]

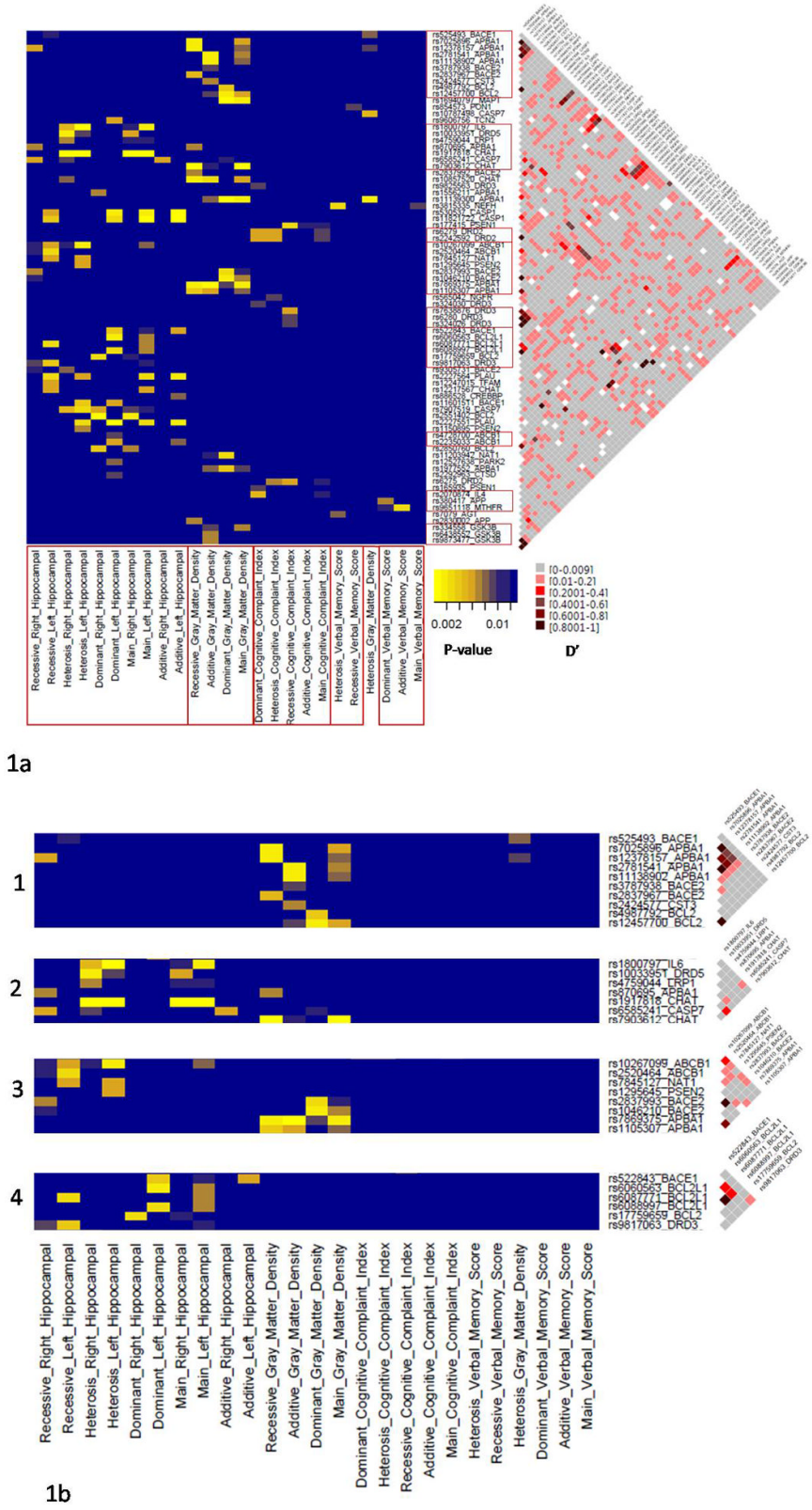

Figure 1.

a- Complete heatmap for all SNPs with at least a single *p*-value below 0.01 (in yellow). All *p*-values above 0.01 are shown in blue. Significant clusters of endpoints and SNPs are boxed in red (verbal memory score all clustered together significantly, as indicated by the connecting red line between the boxes). Alongside the heatmap is an LD plot for the SNPs in the same order as the heatmap. The LD statistic is D' as calculated within *Powermarker*.
b- A close-up of each cluster along with LD plot.

Table 1

Average population (standard deviation) characteristics for major covariates and diagnostic criteria by case/control status and sex.

	Cases (n=85)		Controls (n=47)	
	Male (n=38)	Female (n=47)	Male (n=14)	Female (n=33)
Age	73.1 (5.74)	73.70 (6.51)	71.79 (5.63)	70.88 (4.92)
Education	17.7 (2.66)	15.13 (2.84)	18.21 (1.93)	16.27 (2.40)
APOE (ε4)	n=13	n=18	n=3	n=15
Right Hippocampal	3.3755 (0.51)	3.5098 (0.39)	3.6411 (.040)	3.6736 (0.41)
Left Hippocampal	3.2705 (0.38)	3.3335 (0.39)	3.5085 (0.48)	3.4291 (0.37)
Verbal Memory	−1.3329 (1.22)	−1.3673 (1.26)	0.0003 (0.63)	−0.0272 (0.72)
CCI	31.5300 (13.14)	33.2622 (11.06)	9.5505 (6.04)	9.7607 (6.16)
Gray Matter Density	0.6238 (0.04)	0.6358 (0.03)	0.6473 (0.03)	0.6633 (0.03)

Table 2

The hypothesized role of each gene found in one of the three clusters on AD and MCI based on a literature search with sample references.

GENE (alias)	Suggested role in AD/MCI
ABCB1	Regulates beta-amyloid levels (Kuhnke and Jedlitschky et al. 2007, Lam and Liu et al. 2001)
APBA1 (MINT1, X11)	Binds APP and affects cleavage and translocation(Saito and Sano et al. 2008, Ho and Liu and Sudhof. 2008, Miller and McLoughlin and Lau and Tennant and Rogelj. 2006)(Also known as MINT1, X11)
BACE1	Cleaves APP (Haass. 2004, McConlogue and Buttini et al. 2007, Willem and Lammich and Haass. 2009)
BACE2	Cleaves APP, a BACE1 homolog (Stockley and O'Neill. 2007), increases IL-1R2 secretion (Kuhn and Marjaux and Imhof and De Strooper and Haass and Lichtenthaler. 2007)
BCL2	Induces apoptosis (Lu and Kwong and Li and Wang and Feng and Yew. 2005)
BCL2L1 (BCL-X)	Anti-apoptotic signalling (Lukiw and Bazan. 2006, Shimohama. 2009)
CASP7	Apoptosis regulator, neuron loss in AD (Matsui and Ramasamy et al. 2006, Pompl and Yemul et al. 2003)
CHAT	Synthesizes acetylcholine, which is depleted in AD (Burgess and Saini and Weng and Aubert. 2009), ChAT fibers increasingly immunorreactive in AD and MCI (Cuello and Bruno and Bell. 2007)
CST3	Studies show mixed results, colocalizes with beta-amyloid (Lin and Wang and Wu and Chuo and Kuo. 2003, Monastero and Camarda et al. 2005, Nacmias and Bagnoli et al. 2006)
DRD3	Associated with depression symptoms that co-occur with AD associated (Serretti and Olgiati and De Ronchi. 2007)
DRD5	Connection to AD uncertain, normally functions as dopaminergic receptor (Cosentino and Colombo et al. 2009)
IL6	Inflammatory response, tau phosphorylation (Papassotiropoulos and Hock and Nitsch. 2001, Quintanilla and Orellana and Gonzalez-Billault and Maccioni. 2004)
LRP1	Inolved in APP processing and trafficking (Waldron and Heilig et al. 2008, Yamada and Hashimoto et al. 2008)
NAT1	Folate metabolism (Johnson and Bell and Jonovska and Budge and Sim. 2004)
PSEN2	Gamma-secretase complex formation with PSEN1, well-established AD susceptibility gene (Bertram. 2009, Bertram and Tanzi. 2008, Marcon and Di Fede et al. 2009)

Table 3

Chilibot results for clusters 1–4, showing each gene comparison, the number of interactive and parallel sentences found for each gene pair, and the relationship suggested by the mined literature.

Chilibot Results

	Gene 1	Gene 2	Interactive	Parallel	Suggested Relationship
Cluster 1	BACE1	BACE2	17	24	homologs, amyloid formation
		APBA1	5	9	BACE1 involved in APBA1 generation
		CST3	0	4	implicated in Alzheimer's
	BACE2	APBA1	3	5	BACE2 processes APBA1
		BCL2	2	4	both involved in apoptosis
	APBA1	CST3	17	15	CST3 binds and prevents APBA1 deposition
		BCL2	9	3	APBA1 influences apoptosis through BCL2
	IL6	CASP7	0	3	sometimes upregulated together
Cluster 2		CHAT	7	10	IL6 increases CHAT
		APBA1	4	2	APBA1 upregulates IL6
		LRP1	4	2	implicated in Alzheimer's
	CASP7	APBA1	0	1	implicated in Alzheimer's
	APBA1	LRP1	9	5	APBA1 clearance mediated by LRP1
	CHAT	APBA1	6	4	APBA1 decreases CHAT
Cluster 3	PSEN2	BACE2	0	2	implicated in Alzheimer's
	BACE2	ABCB1	1	0	studied together in myeloma
	ABCB1	NAT1	0	1	studied together in lung cancer
Cluster 4					No relationships found