Automatic Gene Prioritization in Support of the Inflammatory Contribution to Alzheimer's Disease

Stephanie K Furniss, MS[1], Robert Yao[1], Graciela Gonzalez, PhD[1]
[1]Arizona State University, Phoenix, AZ

Abstract

This research seeks to extend the process of novel therapeutic gene target discovery for the treatment of Alzheimer's disease (AD). Gene-gene and gene-pathway annotation tools as well as human analysis are used to explore likely connections between potential gene targets and biochemical mechanisms of AD and associated genes. Rule-based annotation systems, such as GeneRanker, can be applied to the continuously growing volume of literature to extract relevant gene lists. The subsequent challenge is to abstract biological significance from associated genes to aid in discovery of novel therapeutic gene targets. Automatic annotation of genes deemed significant by data-driven assays and knowledge-driven analysis is limited. Therefore, human analysis is still crucial to exploring novel gene targets and new disease models. This research illustrates a method of analysis of an extracted gene list which lead to the discovery of KNG1 as a possible therapeutic target, suggests a connection between inflammation and AD pathogenesis.

Introduction

Alzheimer's disease (AD) is a neurodegenerative disease that causes progressive decline in cognitive functions and impairment of memory. In 2013, AD affects an estimated 5.2 million Americans, and costs of care to the country are projected to reach $203 billion (1). There is no method to prevent, suspend or reverse disease progression; as a result prevalence of AD in the United States is expected to continue its increasing trend. Given its impact on public health, current efforts in AD research are focused on finding as many of the genetic and molecular bases of the disease for the purpose of eventual treatment. Genome-wide association studies (GWAS) have identified novel genetic variants associated with AD (2, 3). But conventional GWAS approaches may not be able to detect some genetic variation or gene-gene interactions (4). This study in particular seeks to extend the process of novel therapeutic gene target discovery.

Published literature is a rich source of genomic information, but the volume of published literature is too large for a biomedical researcher, or group of researchers, to remain up-to-date. Where biomedical researchers may use molecular assays to identify a set of disease-associated genes or genes of interest, biomedical text-mining researchers use integrative gene annotation methods to produce an expanded network of genes associated with a disease. The amount of genes deemed significant from automatic data-driven (large-throughput) assays and knowledge-driven analysis is currently limited. Thus, human analysis is still crucial to exploring novel gene targets and new disease models.

In order to accomplish this, previously manually curated GWAS data related to AD are used as a seed set for the automatic detection of novel and potentially relevant gene targets contained within the literature. Using the results of these experiments, automatic analysis methods more robust than basic statistical methods were employed to extract meaningful genomic relationships. Such work can aid in discovery of novel therapeutic gene targets by informing bench researchers of AD-relevant genes that may be worthy of further investigation or provide evidence to support current hypotheses. It can also be used by biomedical text-mining researchers to guide development of an automated tool that incorporates gene-pathway associations to facilitate understanding of a large gene list, filtering for relevance to research interests, and discovery of novel gene-disease associations via pathway analysis.

Methods

Gene list from AlzGene.org database, which is an actively maintained field synopsis of genetic association studies in Alzheimer's disease, defined the gold standard curated gene list—AlzGene623. AlzGene623 is composed of 623 genes collected from the AlzGene.org database (5). The second seed list named GwasList consists of 547 unique genes from a list of significant SNPs from a GWAS study completed by a collaborating biomedical researcher. The third seed list, SearchByAD, consists of 742 genes and is a product of GeneRanker's gene-disease annotation function.

GeneRanker is an automated gene prioritization tool developed by the Diego lab at Arizona State University to aid biomedical researchers in the discovery of novel gene therapeutic targets. GeneRanker's integrative method examined gene-gene annotations to expand on each seed gene list, resulting in an enriched gene list. It has been evaluated in the context of brain cancer research and atherosclerosis (6, 7). GeneRanker's

computational method combines data extracted from the literature and from curated sources, such as Genetic Association Database and NCBI Gene database.

The four uniquely-developed seed lists were entered into GeneRanker. Each provided a subsequent enriched gene list after gene-gene annotations were examined through an integrative method. Table 1 provides a tabular comparison of the seed lists to show the number of overlapping genes between lists. Combined, the seed lists comprise 1523 unique genes (SeedGenes).

	GwasList	SearchByAD	AlzGene623
GwasList	547	27	25
SearchByAD		742	353
AlzGene623			623

Table 1. Comparison of the three seed lists. Table gives the number of overlapping (shared) genes between two seed lists.

Each enriched gene list was plotted on a computational score versus computational rank graph. A cut off point located where the curve began to flatten was marked (AlzGene623 enriched gene list and cut-off point is graphed in Figure 1). Genes right of the cut off threshold were disregarded due to their low rank and score. Genes to the left of the cut-off point (highly-ranked) may contain potential novel target genes as well as seed genes. For example, of the 990 highly-ranked genes from AlzGene623 expanded gene list 23.2% are in AlzGene623, providing 760 genes for further consideration as novel therapeutic targets. This process was followed for seed lists GwasList and SearchByAD as well resulting in, respectively, 609 potentially novel of the 645 highly-ranked and 711 potentially-novel of the 997 highly-ranked genes. 236 genes were common between these three potentially-novel highly-ranked genes. The 236 extracted potential gene targets (ExPot list) will be further examined using gene-pathway analysis.

GATHER was used for KEGG Pathway (pathway) enrichment analysis of gene lists. GATHER is a tool that integrates various forms of available data and applies a statistical model that quantifies the significance of functional associations (8). It was developed to be used by biomedical researchers to understand the function of a group of genes by showing the user the annotations that distinguish the genes entered from other genes in the genome (8). Enrichment analysis are favorable because biological processes are made up of a group of genes, as opposed to an individual single gene (9). The enriched gene lists were entered into GATHER, which then returned associated gene functions and biological pathways annotated by KEGG Pathway.
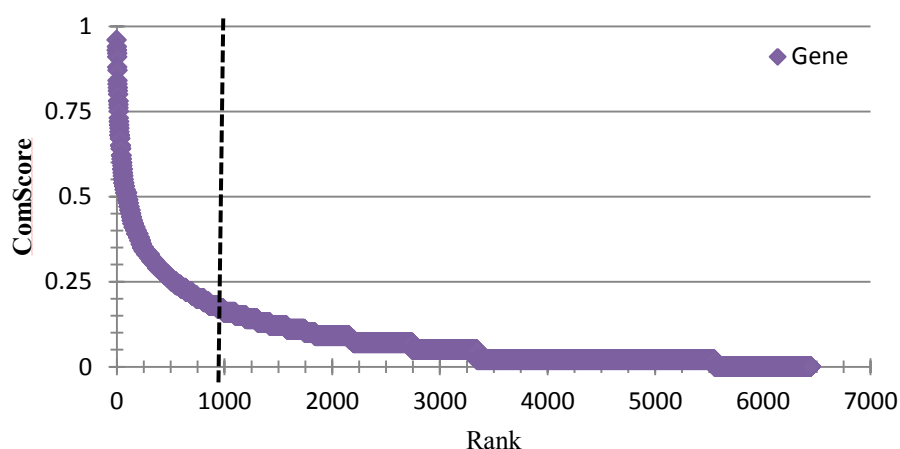
Figure 1. Using GeneRanker, the AlzGene623 seed list expanded to 6449 genes. 990 highly-ranked genes lie to the left of the cut-off point (marked by the dashed line). 23.2% of the highly-ranked genes are in the seed set (AlzGene623), providing 760 genes for further consideration as novel therapeutic targets.

Results

Pathway analysis used significant pathways associated with all seed genes (SeedGenes list), the ExPot list, the combination of the SeedGenes and ExPot (AllGenes list), as well as the four individual seed lists (Table 2). SeedGenes had nine significant (p<.01) pathways (Table 2 a-i). ExPot list had 15 significant pathways (Table 2 c, e, g, j-q, s-v). AllGenes had 18 significant pathways (Table 2 a-g, i-s). Table 2 gives the number of genes per associated pathway and the pathway ranking per gene list.

22 of the 1523 SeedGenes were associated with AD pathway (Table 2 a). The ExPot list had zero genes associated to the AD pathway when examined alone and when examined in combination with SeedGenes, in AllGenes (Table 2 a).

Four of the nine SeedGene pathways ranked lower on addition of the ExPot genes (Table 2 a, b, d, f). Three SeedGene pathways increased their ranking on addition of the ExPot genes (Table 2 c, e, i). One did not change ranking (Table 2 g) and one was not ranked in analysis of AllGenes (Table 2 h).

For two pathways there were ExPot genes associated with the pathway when examined in the context of AllGenes, but not when examined in the context of the ExPot list on its own (Table 2 d, f). In one case, 31 genes in AllGenes are associated with the Complement and coagulation cascades pathway. One of the 31 is in the curated AlzGene623 list (CR1). Five of the 31 genes are in the ExPot list (Table 3). Interestingly, these five genes do have a significant association to the complement and coagulation cascades pathway when examined in context of their source gene list, ExPot list.

KEGG Pathway	SeedGenes	AllGenes	ExPot	GwasList	AlzGene623	SearchByDisease
a Alzheimer's disease	22 (1)	22 (3)	--	4 (5)	20 (1)	21 (1)
b Pyrimidine metabolism	2 (2)	2 (5)	--	--	1 (6)	--
c Insulin signaling pathway	46 (3)	74 (2)	28 (3)	--	29 (2)	--
d Neuroactive ligand-receptor interaction	79 (4)	99 (13)	--	19 (4)	--	--
e Apoptosis	33 (5)	49 (4)	16 (10)	--	19 (4)	22 (7)
f Complement and coagulation cascades	26 (6)	31 (16)	--	--	14 (7)	23 (3)
g Calcium signaling pathway	54 (7)	75 (7)	21 (14)	15 (2)	--	--
h Prostaglandin and leukotriene metabolism	16 (8)	--	--	--	--	13 (5)
i Purine metabolism	13 (9)	13 (8)	--	--	3 (3)	1 (2)

Table 2. Subset of the full pathway analysis results from GATHER for seven gene lists examined. The ranking of the pathway is shown in parentheses for each gene list.

Discussion

Three seed lists identified from three unique sources were applied to an automated gene enrichment and prioritization tool to expand on genes already known for the discovery of novel genes related to AD. The overlap of these expanded gene lists were used to identify potential gene therapeutic targets, ExPot list. The ExPot list was further narrowed using pathway analysis of relevant genes. The findings of this study confirm a known association to AD while supporting the potential of the ExPot genes. They also demonstrate an approach to gene- and pathway-disease analysis can lead to a single gene or group of similar genes worthy of further evaluation.

An important, though not surprising, finding with respect to AD is that the pathway analysis of SeedGenes ranked AD as the most significant pathway (Table 2, a) confirming the SeedGenes association to AD. Further, no ExPot gene is associated to the AD pathway and the addition of the 236 ExPot genes dilutes the ranking of the AD pathway (Table 2, a) suggesting confirmation that a novel gene-AD association may lie within the ExPot genes.

For two pathways—complement and coagulation cascade and neuroactive ligand-receptor pathway—there were ExPot genes associated with each pathway when the list was examined in context of AllGenes, but not when

examined in the context of the ExPot list on its own. One of these, the complement and coagulation cascade pathway, had five ExPot genes associated to it (Table 3). This could suggest that the five genes are well-studied and have a weak correlation to the pathway or it could suggest the genes are not well studied, which, if true, increases the likelihood that they too have not been examined in relation to AD. The complement cascade is an indispensable element of the innate immune response (10) and neuroinflammation is believed to be an underlying mechanism in AD, therefore the complement cascade is relevant to AD pathogenesis.

Of the five ExPot genes implicated in the complement and coagulation cascades (Table 3), four are receptors and one, kininogen 1 (KNG1), is a cofactor to coagulation and inflammation. PubMed literature search revealed that a gene-AD association has been examined for some of these ExPot genes (Table 3), consequently removing them from consideration as a potential novel AD therapeutic target.   Coagulation factor II (thrombin) receptor (F2R) is not further explored because the lack of evidence connecting AD and coagulation. KNG1 encodes high molecular weight kiniongen protein (HMWK), which plays a role in inflammation, regulation of blood pressure, and coagulation; therefore KNG1 and its precursors are further examined here. As an immune-cell membrane protein, complement component (3d/Epstein Barr virus) receptor 2 (CR2, CD21), is involved in immune responses.

Gene Symbol	Gene Name	Pathway	Examined association with AD?	Citation(s)
BDKRB2	bradykinin receptor B2	Co, Ca, N, Cy	Yes	(11), (12)
C5AR1	complement component 5a receptor 1	Co, N	Yes	(13)
CR2	complement component (3d/Epstein Barr virus) receptor 2	Co	No	
F2R	coagulation factor II (thrombin) receptor	Co, Ca, N	No	
KNG1	kininogen 1	Co	No	
Pathway key: "Co" is Complement and coagulation cascades; "Ca" is Ca^{2+} signaling pathway; "N" is Neuroactive ligand-receptor interaction; "Cy" is Regulation of actin cytoskeleton.				

Table 3. Five ExPot genes associated with the Complement and coagulation cascade pathway.

Cleavage of HMWK results in bradykinin (BK) and cleaved high-molecular-weight kininogen (HKa). HKa, using intracellular signaling pathways, contributes to the pathogenesis of inflammatory diseases by releasing cytokines TNF-α, IL-1β, IL-6, and chemokines IL-8 and MCP-1 from isolated human mononuclear cells (14). HKa may exert antiadhesive effects, thereby regulating leukocyte recruitment into inflamed tissue (15). TNF-α, IL-1β, IL-8, and IL-6 are in the curated gene list therefore providing a strong relationship to genes that have known association to AD.

BK acts through receptors BDKRB2 (ExPot list) and BDKRB1 (16) to mediate activation of proinflammatory signals and regulate cardiovascular processes. A recent study suggested activation of BDKRB1 as a novel therapeutic approach for AD based on evidence that BDKRB1 activation plays an important role in limiting the accumulation of Aβ in AD-like brain possibly through the regulation of activated glial cell accumulation and release of pro-inflammatory mediators (16). No human-model studies of BDKRB1 in association to AD were found in a PubMed search.

As part of the inflammatory response, the complement cascade and KNG1 may deserve greater attention in the search for therapeutic targets for the treatment of AD. In fact, it has been suggested that antibodies to kininogen or peptidomimetics might be a useful and safe therapy in inflammatory diseases or sepsis involving cytokines (17).

Conclusion

Examination of the five ExPot genes associated with the complement and coagulation cascade lead to the novel identification of KNG1 as a potential therapeutic target for treatment of AD. Previous research suggesting that antibodies to kininogen might be a useful and safe therapy in inflammatory diseases involving cytokines (17) further increases the interest and likelihood that KNG1 could be a therapeutic target for treatment of AD.

This research again confirmed GeneRanker, while it also created credibility of the prioritization-extracted gene list through manual review of published scientific literature and automatic annotations.

It is possible that the highly-ranked extracted gene lists include false positives, genes that have been previously studied in association with Alzheimer's but were not in the seed lists or 'noisy' genes. It is important to consider such false positives to examine how our biological knowledge base is driving the gene extraction. For

example, there may be noise from a large variation. As a result, the gene annotation method reaches genes which encode numerous protein kinases and MAP kinases, which are not disease specific.

In this study, pathway analysis was applied to relevant GWAS data (GwasList) in an effort to weed through the noise and narrow down potential gene target lists. Our lab is exploring methods for identifying 'noisy' genes

A gene-enrichment study relies on pathway analysis that is only as good as the functional information providing its pathway definitions. The differences across pathway databases can lead to divergent enrichment analysis results tool to tool. Only one tool, GATHER, was used for gene-enrichment pathways analysis in this study. It may be advisable to incorporate pathway analysis results from another tool, such as DAVID, as well.

Acknowledgements

References

1.   Alzheimer's Association. 2013 Alzheimer's disease facts and figures. Alzheimers Dement. 2013;9(2):208-45.

2.   Harold D, Abraham R, Hollingworth P, et al. Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. Nat Genet. 2009 Oct;41(10):1088-93.

3.   Lambert J, Heath S, Even G, et al. Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. Nat Genet. 2009 Oct;41(10):1044-9.

4.   Morgan K. The three new pathways leading to Alzheimer's disease. Neuropathol Appl Neurobiol. 2011 Jun;37(4):353-7.

5.   Bertram L, McQueen M, Mullin K, Blacker D, Tanzi R. Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database. Nat Genet. 2007 Jan;39(1):17-23.

6.   Gonzalez G, Uribe JC, Armstrong B, McDonough W, Berens ME. GeneRanker: An Online System for Predicting Gene-Disease Associations for Translational Research. Summit on Translat Bioinforma. 2008;26:5.

7.   Gonzalez G, Uribe JC, Tari L, Brophy C, Baral C, editors. Mining Gene-Disease Relationships from Biomedical Literature: Weighting Protein-Protein Interactions and Connectivity Measures. Pac Symp Biocomput; 2007; Maui, Hawaii.

8.   Chang JT, Nevins JR. GATHER: a systems approach to interpreting genomic signatures. Bioinformatics. 2006;22(23):2926-33.

9.   Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 2009 Jan 1;37(1):1-13.

10.   Aiyaz M, Lupton MK, Proitsi P, Powell JF, Lovestone S. Complement activation as a biomarker for Alzheimer's disease. Immunobiology. 2012 Feb;217:204-15.

11.   Prediger RDS, Medeiros R, Pandolfo P, et al. Genetic deletion or antagonism of kinin B1 and B2 receptors improves cognitive deficits in a mouse model of Alzheimer's disease. Neuroscience. 2008;151(3):631-43.

12.   Mendonsa G, Dobrowolska J, Lin A, Vijairania P, Jong YJ, Baenziger NL. Molecular Profiling Reveals Diversity of Stress Signal Transduction Cascades in Highly Penetrant Alzheimer's Disease Human Skin Fibroblasts. PLoS One. 2009;4(2):e4655.

13.   Ager RR, Fonseca MI, Chu SH, et al. Microglial C5aR (CD88) expression correlates with amyloid-beta deposition in murine models of Alzheimer's disease. J Neurochem. 2010;113(2):389-401.

14.   Khan MM, Bradford HN, Isordia-Salas I, et al. High-Molecular-Weight Kininogen Fragments Stimulate the Secretion of Cytokines and Chemokines Through uPAR, Mac-1, and gC1qR in Monocytes. Arterioscler Thromb Vasc Biol. 2006 October 1, 2006;26(10):2260-6.

15.	Chavakis T, Kanse SM, Pixley RA, et al. Regulation of leukocyte recruitment by polypeptides derived from high molecular weight kininogen. FASEB J. 2001 November 1, 2001;15(13):2365-76.

16.	Passos GF, Medeiros R, Cheng D, Vasilevko V, LaFerla FM, Cribbs DH. The Bradykinin B1 Receptor Regulates Aβ Deposition and Neuroinflammation in Tg-SwDI Mice. Am J Pathol. 2013;182(5):1740-9.

17.	Khan M, Liu Y, Khan M, et al. Upregulation of tissue factor in monocytes by cleaved high molecular weight kininogen is dependent on TNF-alpha and IL-1beta. Am J Physiol Heart Circ Physiol. 2010 2010 Feb;298(2):H652-8.