

**Research Article** 

# An Exploratory Analysis of Conservation of Co-Expressed Genes across Alzheimer's disease Progression

Pradeep Chowriappa\*, Prerna Dua and Walter J Lukiw

Department of Computer Science, 241 Nethken Hall, Louisiana Tech University, Ruston, LA - 71272, USA

### Abstract

Alzheimer's disease (AD) is a complex disease where the analysis of gene expression patterns relies on computational techniques to understand the cause and progression of the disease. Evidence postulates that the complexity of AD stems from the overlap of early-stage markers with normal aging. Furthermore, there is increasing evidence suggesting that gene co-regulation in AD plays a vital role in the progression of the disease. The aim of this work is to identify and track co-regulated genes from incipient to severe cases of AD i.e. samples that exhibited progression of AD. We hypothesize that co-expressed genes associated with two markers of AD (the cognitive marker-mini mental state examination (MMSE) and the pathological marker Neurofibrillary Tangles (NFT)), are conserved across AD progression. For our analysis we used the Blalock dataset and the prominent tool weighted correlation network analysis (WGCNA). Through our analysis we observed that genes GNA11 and MAP2K2 were consistently ranked through the progression of AD. The functional analysis of the identified co-regulated genes at the incipient to severe stages of AD includes RNA and cofactor binding. Through this exploratory study we conclude that from incipient to severe stages of AD the gamut of co-regulated genes vary rather than being conserved across disease severity.

**Keywords:** Alzheimer's disease; Hippocampus; Gene co-expression networks; Data Mining; Microarray; Gene expression

# Introduction

Alzheimer's is a complex disease (AD) progressive neurodegenerative disorder of the brain. The complexity of this disease has been attributed to several pathogenic factors that initiate the processes associated with AD. These pathogenic factors, are still elusive, are responsible for processes such as abnormal  $\beta$  amyloid (A $\beta$ ) production [1], tau pathology and the progressive formation of neurofibrillary tangles [2], inflammation, synaptic loss, oxidative damage, protein processing or misfolding [3], calcium dyshomeostasis [4], aberrant reentry of neurons into cell cycle, cholesterol synthesis and effects of hormones, or growth factors. This illusive nature of AD has been attributed to three factors, namely (a) the complexity of the disease itself, (b) the variability of gene expression resulting from the cellular heterogeneity of brain tissues, and (c) from interindividual genetic variability [5]. Blalock et al. [6] in conjunction with microarray analysis investigated the role of genes in discriminating between subjects that had mild cognitive impairment, incipient AD, and those with normal aging [6]. They successfully linked patterns of gene expression to a cognitive marker, mini-mental status examination (MMSE) and a pathological marker neurofibrillary tangle (NFT) of AD, independent of AD diagnosis. Furthermore, they reported that these markers are negatively correlated and tested this correlation with the gene expression patterns of 31 separate microarray samples [6]. They identified genes with potentially important implications for the early pathogenesis of AD specifically in the incipient stage of the disease. This observation of gene expression correlating to the markers was further reinforced by Dunckley et al. [5].

In recent times, there is a growing interest in using co-expression analysis to identify subsets of co-expressed genes that act as molecular signatures for genetic disease. Gene co-expression as defined by Lee et al. [7] is an approach in which genes that have similar expression patterns across a set of samples are hypothesized to either have a or belong to a common functional process [8,9]. As shown by Ray and Zhang [10], used gene co-expression analysis to identify the differences among regions of the brain to shed light onto the progression of AD. Prominent computational techniques used for the identification of co-expressed gene subsets include bi-clustering [11]. Bi-clustering techniques are aimed at finding co-expressed genes across a set of microarray samples. However, bi-clustering techniques rely solely on gene expression patterns across samples to establish function similarity. More recently, techniques of co-expression gene networks have become a popular technique used to identify functional relationships between co-expressed genes. Based on the principles of graph theory, these coexpression gene network techniques uses both gene expression patterns and information from prominent markers to identify closely connected groups of genes that represent co-expressed genes. According to Ray et al. [12], gene co-expression network analysis was successfully used to establish the relation between functional clusters of genes and cisregulatory elements to identify differentially expressed genes used to elucidate the biological processes involved in AD.

Based on the observations derived from Blalock et al. [6], the objective of this study is to identify the co-expressed genes in the incipient AD hippocampus samples using the cognitive marker MMSE and the pathological marker NFT. We hypothesize that co-expressed genes from the incipient AD samples are conserved across different degrees of AD severity and this conservation is best captured by taking into consideration the negative correlation between markers of MMSE and NFT. We therefore aim to create a functional map of these incipient

\*Corresponding author: Pradeep C., Department of Computer Science, 241 Nethken Hall, Louisiana Tech University, Ruston, LA - 71272, USA, Tel: +001-31-82574612, Fax: +001-31-82574922; E-mail: pradeep@latech.edu

Received August 07, 2013; Accepted August 26, 2013; Published August 31, 2013

Citation: Pradeep C, Prema D, Lukiw WJ (2013) An Exploratory Analysis of Conservation of Co-Expressed Genes across Alzheimer's disease Progression. J Comput Sci Syst Biol 6: 215-227. doi:10.4172/jcsb.1000119

**Copyright:** © 2013 Pradeep C, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

co-expressed genes and track their evolution through the progression of the disease.

We focus our discussion on the schematic representation of the proposed methodology and the steps entailed. In the following sections we describe the results obtained and the function interpretation of the gene clusters identified using the gene ontology enrichment analysis software toolkit (GOEAST)<sup>1</sup>\ [13] DAVID-EASE<sup>2</sup> [14], and the tool for pathway analysis known as Pathway Express<sup>3</sup> [15]

### Materials and Methods

The proposed methodology was developed based on the data used by Blalock et al. [6]. The data set consists of expression profiles of the brains hippocampi. Obtained from 22 postmortem subjects with Alzheimer's disease (AD) at various stages of severity, the dataset consists of seven incipient AD samples, eight moderate AD samples, and seven severe AD samples respectively. The dataset also contains nine control samples, resulting in a dataset consisting of 31 samples. Each sample is scored by both (a) cognitive marker MMSE and (b) the pathological marker NFT.

Related studies on AD first select a set of differentially expressed genes on which further analysis is performed [10,16,17]. However, comparing lists of genes from various AD studies has been shown to be computationally inefficient as the resultant computational techniques developed become gene specific. Keeping this in mind we propose a framework that is aimed at providing a comprehensive view of the coexpression patterns between genes using the technique of weighted co-expression network analysis (WGCNA) [8,17]. Our objective is to, organizing co-expressed genes into gene clusters (aka modules) that are believed to be of functional significance. The functional significance of genes in each module is determined using both (a) similarity between gene expression profiles and (b) maximum average negative correlation between the cognitive trait MMSE and pathological trait NFT respectively. The proposed framework therefore consists of twostages (as shown in Figure 6) that encapsulates (a) the identification and ranking gene modules/clusters based on significance, and (b) the identification of conserved co-expressed genes through the disease progression.

Lastly, it will be further interesting to use these types of algorithms in the analysis of larger sample sets of incipient, moderate, and severe AD to further understand what genetic networks are selectively targeted in the early, moderate and late stages of the AD process, and what novel therapeutic strategies may be then devised for the clinical management of this expanding health care concern.

#### Identification and ranking of gene clusters / modules

As shown in Figure 6, before describing the process of identification and ranking of gene co-expression gene modules/clusters, the gene expression data is subjected to data preprocessing strategy employed.

**Data preprocessing:** As described, the dataset consisted of 31 samples. Each sample consisted of 22,283 genes. These genes were subjected gene filtering, where genes with p-value<0.5 were considered to be significant. We further pruned the number of genes to find those genes that were common across all thirty one samples in the dataset. The resultant gene set consisted of 4,483 genes. The resultant

1 http://omicslab.genetics.ac.cn/GOEAST/index.php

2 http://david.abcc.ncifcrf.gov/

gene set was standardized using the z-score normalization, where the gene expression values were normalized to have a common mean and standard deviation.

WGCNA Analysis: The proposed framework consists of a one versus one (pair-wise) comparison of sample classes, categorized by the degree of Alzheimer's disease severity. In the case of control versus incipient AD (CI) we considered the normalized expression data of nine control samples along with the seven incipient samples for analysis.

#### Identification of co-expressed genes

To effectively identify clusters of co-expressed genes for each of the combination of classes, we used the method weighted gene coexpression network analysis (WGCNA) [18,19]. WGCNA treats each gene (i.e., gene expression profile across a set of samples) as node in a network. The *i*<sup>th</sup> gene expression profile  $x_i$  is a vector whose components report the gene expression values across *m* samples. We used the co-expression similarity  $s_{ij}$  as the absolute value of the correlation coefficient between their expression profiles between two genes *i* and *j*.

$$s_{ij} = \left| cor(x_i, x_j) \right|, \tag{1}$$

As prescribed in WGCNA, we used the concept of soft threshold to equate the co-expression similarity  $s_{ij}$  into a measure of connection strength between genes *i* and *j*, respectively. This transformation was brought about by considering the power of the absolute value of the correlation coefficient represented as

$$a_{ij} = s_{ij}^{\beta}$$
, where  $\beta \ge 1$ . (2)

The choice of the threshold  $\beta$  controls the sensitivity and specificity of the pairwise connection strength independent of the size of the network and the significance of correlation between expression profiles of a pair of genes. Co-expression networks aim at detecting subsets of genes (gene clusters) that are tightly connected to each other. We used topological overlap (TO) as measure of dissimilarity to identify biologically significant cluster of genes [20]. The TO ( $\omega_i$ ) of gene *i* with other genes in a network is represented as

$$p_i = \sum_{i}^{n} \omega_{ij} , \qquad (3)$$

Thus, a gene *i* with a high  $\omega_i$  has a high overlap with many other genes in the network. The topological overlap of two nodes reflects their relative interconnectedness. The topological overlap matrix (TOM)  $\Omega = [\omega_i]$  provides a similarity measure between multiple genes in the gene co-expression network represented as

$$\omega_{ij} = \frac{l_{ij} + a_{ij}}{\min\{k_i, k_j\} + 1 - a_{ij}},$$
(4)

where  $l_{ij} = \sum a_{iu}a_{uj}$ , and  $k_i = \sum a_{iu}$  is the node connectivity. As  $\omega_{ij}$  is a measure of similarity it was subtracted from one, i.e. to convert it to a measure of dissimilarity

 $\Psi_{ij}^{\omega} = 1 - \omega_{ij}. \tag{5}$ 

In this work we used the TOM-based dissimilarity  $\Psi_{ij}^{\omega}$  to cluster gene expression profiles for module identification. It is believed that TOM-based dissimilarity  $\Psi_{ij}^{\omega}$  leads to more distinct cluster of genes.

#### Identification of gene clusters/modules

We define a cluster of genes (module) whose expression profiles are highly correlated across the samples and with high topological overlap. We used the average linkage hierarchical clustering coupled with the TOM-based dissimilarity  $\Psi_{ij}^{\omega}$ , to cluster genes that exhibit

<sup>3</sup> http://vortex.cs.wayne.edu/projects.htm#Pathway-Express

coherent expression profiles. This clustering was implemented using the dynamic tree cut library of the WGCNA package [21] to identify coherent clusters of genes.

#### Cluster ranking and selection

Once the clusters of co-expressed genes were identified, our immediate objective was to choose those clusters that are biologically significant. In this paper we define the significance (GS) of a cluster of genes based on the averaged values of the markers MMSE and NFT scores across all genes in the cluster [6].

In determining the GS of a cluster of genes, we assign the nonnegative marker score to each gene (assigned by correlation between genes to sample) and averaged across all the genes in that cluster. In this work we deal two traits - MMSE and NFT. We therefore carry out the ranking of clusters independently for each trait. Based on our hypothesis, we believe that clusters of genes that exhibit the higher average deviation of the GS between traits MMSE and NFT relate to higher functional significance. The average deviation  $\delta$  for a cluster of genes M between the GS scores of MMSE and NFT is captured using the following relation.

 $\delta(M) = \frac{1}{2} \sum_{i=1}^{2} |x_i - \overline{x}|, \qquad (6)$ 

## Functional significance and conservation of genes

As previously stated, the above approach is carried out on three (one versus one – pair wise) combinations that include severe AD versus incipient AD (SI), control versus incipient AD (CI), and moderate AD versus incipient AD (MI). For each of the above combinations we extract the top four ranked clusters of genes based on average deviation of GS (based on equation 6).

In order to determine the molecular function of a set of genes commonly expressed across AD severity, we take into consideration the pair wise combinations CI, MI, and SI, to identify subsets of genes that are commonly expressed across AD severity (Figure 7). Since these combinations share a common sample set (incipient AD), we believe that we should observe a set of commonly expressed genes.

#### Functional annotation of gene clusters

Database for Annotation, Visualization and Integrated Discovery (DAVID) is a web-based tool that provides integrated solutions for the annotation and analysis of genome-scale datasets derived from highthroughput technologies such as microarrays. Analyses of results are dynamically linked to primary data and external data repositories, to provide an in-depth as well as broad-based data coverage. DAVID EASE is useful for summarizing the predominant biological "theme" of a given gene list. Given a list of genes resulting from our analysis is submitted to DAVID and it rapidly calculates over-representation statistics for every possible Gene Ontology term with respect to all genes represented in the data set.

We used DAVID version 6.7 [14] to identify annotation terms significantly enriched in each reference gene set. We used the modified Fisher's exact test, or EASE score, to identify enriched annotation terms derived from GNF-U133A- QUARTILE and gene ontology (GO) annotation terms, which includes biological process (BP), molecular function (MF), and cellular component (CC) categories. We used the more specific GO term categories provided by DAVID, called GO FAT, to minimize the redundancy of the more general GO terms in the analysis to increase the specificity of the terms.

A list of gene symbols was generated for each dataset and used as input into DAVID. We used the Functional Annotation Tool, with the Human Genome U133A plus 2.0 Array as the gene background, to independently analyze each gene set. We used a count threshold of 5 and the default value of 0.1 for the EASE score settings. We also used the Benjamin corrected p-value, with p<0.05 as the significance threshold. Significant annotation terms identified in the GNF annotation category were further filtered using the inter quartile range of the category size, where the 1<sup>st</sup> and 3<sup>rd</sup> quartile were removed from the results. Significant annotation terms with a category size less than 100 and greater than 1000.

Apart for using the tool DAVID EASE for gene enrichment, we used the gene enrichment tool Gene Ontology Enrichment Analysis Software Toolkit (GOEAST) [13]. This tool is predominantly used to identify statistically over-represented GO terms within sets of genes. In our analysis, we subject the probe-ID's of different genes in the gene clusters to generate enriched GO terms in graphical format according to their relationships in the hierarchical tree of each GO category (Biological process, Cellular component, and Molecular function), for a visual representation of relationships between genes.

Similarly to identify the pathways associated with genes of interest, we used the tool Pathway express [15] that is part of the popular tool Onto-express [22].

## Results

#### Co-expressed genes in Incipient Alzheimer's disease samples

To extract a candidate set of genes that are co-expressed in the incipient samples, we use the R package weighted gene co-expression network analysis (WGCNA) [8]. In the experimental setup we consider the seven incipient AD samples from the Blalock dataset. To retain an effective baseline for gene expression we take into consideration the nine control samples throughout this analysis. These samples were preprocessed and normalized resulting in a set of 4,483 genes. We then apply the procedure of WGCNA to identify those genes that are co-expressed in the incipient samples.

As part of WGCNA we used soft threshold ( $\beta$ ) using Pearsons correlation. As shown in Figure 1 the value of the soft threshold  $\beta$  is determined be of power equal to eight, where we observe stability in both the scale free topology and the mean connectivity respectively.

#### **Clusters of genes**

With the soft threshold determined and fixed at eight, in the next step we work towards the identification of clusters of genes. Using the cognitive marker MMSE and the pathological marker NFT independently, we cluster genes based on their topological overlap (TO) using WGCNA. Figure 2 provides the hierarchical clustering of genes based on their TO profiles. We used the automatic module detection using dynamic tree cutting (DTC) to identify clusters of genes.

With the identification of independent clusters, we rank the clusters on the average deviation of MMSE and NFT scores. Based on the observation by Blalock et al. that MMSE and NFT are negatively correlated, we assign the cluster of genes that has the highest average deviation of MMSE and NFT the highest rank ( $M_1$ ). Table 1 shows the ranking of gene clusters in the control versus incipient samples.

A visual representation of the normalized expression profiles of the







genes performed on (A) cognitive marker MMSE and (B) pathological marker NFT based on both control and incipient samples.

genes of the highest ranked gene cluster CI-Purple  $(M_i)$  is shown in Figure 3. The genes of this cluster (Table 2) exhibit maximum average deviation considering the cognitive trait MMSE and the pathological trait NFT. There were 101 genes selected in gene cluster CI-Purple  $(M_i)$ , Figure 3(a) shows the 101 genes expressed in the control samples. Similarly Figure 3(b) shows the expression profile of the same 101 genes in the incipient AD samples. We observe there is a consistency in the expression profiles of the control and incipient genes exhibiting co-expression across sample types.

## Functional profile of selected genes

To ascertain common biological functions associated with the 101 genes of cluster CI-Purple  $(M_i)$ , we adopted an integrated bioinformatics approach based on structured biological knowledge provided by the Gene Ontology (GO) consortium [23]. We applied the GO enrichment analysis software toolkit (GOEAST) [13] to the 101 genes. Briefly, all three branches of GO knowledge structure (biological process (BP), molecular function (MF), and cellular component (CC)) were utilized for this analysis. To test for GO category enrichment, we used the following filtering criteria: (a) p-value cutoff of  $\leq 0.001$ , (b) limited the GO annotation category size of  $\times \leq 1000$  to minimize artificial elevation of p-value, and (c) gene count of >4 in significant categories. Applying these filters, a total of nine enriched GO categories were identified for the 101 genes in the cluster CI-Purple  $(M_i)$ : five BP categories, 10 CC categories, and nine MF categories (Table 3).

For example of enriched categories with the highest gene content per GO branch includes cellular process (BP: 72 genes; p-value= $5.95 \times 10^{-6}$ ); cellular metabolic process (BP: 45; p-value= $1.25 \times 10^{-4}$ ), glucose catabolic process (BP: 6, p-value= $6.68 \times 10^{-5}$ ), ribonucleoprotein complex (CC: 17, p-value= $2.36 \times 10^{-7}$ ), and glyceraldehyde 3 phosphate dehydrogenase (NAD+) (phosphorylating) activity (MF: 5, p-value= $1.38 \times 10^{-9}$ ).

For additional support, we also performed GO enrichment analysis

Gene Cluster (Rank)	MMSE significance	NFT significance	Avg. Deviation	No.of genes
Purple (M <sub>1</sub> )	0.1371608	0.642366	0.2526026	101
Black (M <sub>2</sub> )	0.1866539	0.4573602	0.1353532	136
Red (M <sub>3</sub> )	0.1428014	0.3758707	0.1165347	167
Salmon (M <sub>4</sub> )	0.1669128	0.3841604	0.1086238	62
White	0.1640657	0.3353733	0.0856538	478
Cyan	0.321479	0.1753377	0.0730707	52
Brown	0.143615	0.244084	0.0502345	369
Grey	0.2274596	0.1396617	0.043899	2
Turquoise	0.2119453	0.148309	0.0318182	1550
Tan	0.2411942	0.1786841	0.031255	67
Pink	0.1497869	0.1975164	0.0238648	350
Green Yellow	0.173846	0.2215524	0.0238532	69
Midnight Blue	0.2792168	0.3082386	0.0145109	32
Dark Red	0.1640868	0.1861055	0.0110094	184
Light green	0.146697	0.1553981	0.0043506	327
Dark Grey	0.1565169	0.1637953	0.0036392	304
Light yellow	0.1936461	0.1890657	0.0022902	233

 Table 1: Clusters of genes ranked based on average deviation of MMSE and NFT significance when control and incipient samples are analyzed.





Figure 2: The expression	profiles of co expressed	annes of the gone cluste	$r \cap Purple (M)$	$across(\Lambda)$ control same	les and (R) incinient camples
rigure 5: The expression	promes or co-expressed	i genes of the gene cluste	$(NI_{1})$	across (A) control same	nes, and (B) incipient samples.

LOC644101/ CBX3	NME2 / NME1-NME2 / NME1	PGM1	RAB11FIP3	RPL13 / RPL13P12
RALY	RER1	RALGPS1	DUT	FASN
SMARCC2	LOC407835 / MAP2K2	PCDHGA12 / PCDHGC5 / PCDHGC4 / PCDHGC3	PSMC5	SETD2
COPB1	SYF2	LHX2	NOL7	CDC34
USP7	RANBP9	MT1H	MT1F	LOC100133042 / GAPDH / GAPDHL6
ATP6V1F	LEPROTL1	MAEA	NUDT3	TAOK3
MT1P2	RPL13 / RPL13P12	KAZ	DPP8	GNA11
MAN1A2	MT1F	GLTSCR2 / LOC440311	MIR21	ACTB
CTBP1	SFRS11	POLR2E	LIN7C	CTBP2
SIK3	SLC36A2	CUEDC2	CNOT2	JTB
ZC3H14	RPL13 / RPL13P12	ZSCAN18	SAFB2	ILF2
CEP350	CUX1	NENF	AGPAT1	HSPA9
KRT10	FMR1	ZMAT3	DMPK	NFE2L1
CHN2	MT1L / MT1E / MT1JP / MT1P3	RPL36P14 / RPL36	GNA11	ACTB
MZF1	UBA1	PACSIN2	SFRS8	ZRSR2
EIF1AX	AZIN1	SIRPA	SPTAN1	MTA1
DCTN1	DDX42	RPS2P17 / RPS2 / RPS2P55 / RPS2P8 / RPS2P20 / RPS2P11 / RPS2P12 / RPS2P5 / RPS2P51	APLP2	CDV3
WDR82	FBXO11	NDUFV1	LSM14A	ZRSR2

**Table 2**: The 101 genes that belong to cluster CI-Purple ( $M_{\gamma}$ ).

Category	GO ID	Term	No. of genes in Purple (M1)	No. of Gene Chips	Log-odds-ratio	P-value
ВР	GO:0009987	cellular process	72	23657	0.720328	5.95E-06
	GO:0044237	cellular metabolic process	45	12004	1.021008	0.000125
	GO:0006007	glucose catabolic process	6	131	4.631922	6.68E-05
	GO:0006096	glycolysis	6	98	5.050635	1.33E-05
	GO:0006414	translational elongation	8	267	4.019686	1.62E-05
сс	GO:0005623	cell	92	29177	0.7714	2.72E-13
	GO:0044464	cell part	92	29176	0.77145	2.72E-13
	GO:0005622	intracellular	88	25774	0.886185	8.19E-14
	GO:0044424	intracellular part	87	25017	0.912705	7.77E-14
	GO:0043226	organelle	68	21406	0.782117	5.28E-06
	GO:0043229	intracellular organelle	68	21374	0.784276	5.21E-06
	GO:0043228	non-membrane-bounded organelle	34	6767	1.443544	1.02E-05
	GO:0043232	intracellular non-membrane-bounded organelle	34	6767	1.443544	1.02E-05
	GO:0030529	ribonucleoprotein complex	17	1294	2.830223	2.36E-07
	GO:0022626	cytosolic ribosome	8	200	4.436526	2.85E-06
MF	GO:0004365	glyceraldehyde-3-phosphate dehydrogenase (NAD+)(phosphorylating) activity	5	8	8.40231	1.38E-09
	GO:0016620	oxidoreductase activity, acting on the aldehyde or oxo group of donors, NAD or NADP as acceptor	5	56	5.594955	2.70E-05
	GO:0046870	cadmium ion binding	5	20	7.080382	2.50E-07
	GO:0097159	organic cyclic compound binding	51	12179	1.180699	3.53E-07
	GO:1901363	heterocyclic compound binding	51	12104	1.189611	3.02E-07
	GO:0051287	NAD binding	8	128	5.080382	1.24E-07
	GO:0003723	RNA binding	19	2196	2.227647	9.78E-06
	GO:0005488	binding	83	27702	0.697719	4.82E-08
	GO:0005515	protein binding	64	19718	0.813157	1.02E-05

Table 3: Function profile of the 101 genes of Purple (M1): Enrichment Gene Ontology (GO) categories.

using the DAVID bioinformatics resource, which employs the Fisher's Exact test. With a p-value cut-off of p<0.05, we derived a total of 15 enriched GO categories using DAVID functional analysis: One BP category, 12 CC categories, and two MF categories respectively. While all categories from both analyses related to similar terms, the MF RNA binding matched in both analyses. Refer Table 4 for the different categories derived using DAVID analysis.

We further characterize the functionality of the 101 genes in the CI-Purple  $(M_i)$  cluster through pathway analysis using the tool Pathway Express [5]. Using the KEGG database, there were 45 pathways that correspond to the genes in CI-Purple  $(M_i)$  (Table 5). The significance of each molecular pathway with regards to the 101 genes in CI-Purple  $(M_i)$ is based on the impact factor. We filtered out those molecular pathways that were below impact factor <1.75. The resultant were 17 molecular pathways, of which the top five were the Chronic myeloid leukemia (3 genes, impact factor 5.424), Ribosome (3 genes, impact factor 5.243), Notch signaling pathway (2 genes, impact factor 4.008), Vibrio cholera infection (2 genes, impact factor 3.552), and Huntington's disease (3 genes, impact factor 3.228). Of the 17 significant molecular pathways we also observe the Parkinson's disease pathway and the Alzheimer's disease pathways.

To visualize the relationship among enriched GO categories, we

generated directed acyclic graphs based on GO knowledge base using the tool GOEAST. We focus our attention to the MF graph (Figure 4), enriched terminal nodes relate to RNA binding and nucleotide binding functions of the GO.

We next perform our analysis of conservation of different functional categories through the progress of AD. Here we repeat our analysis of identifying co-expressed genes using samples of moderate AD with incipient AD to extract clusters of genes. Here we observe a cluster of 177 genes MI-Black  $(M_1)$  possessed the highest average deviation of 0.226 between the cognitive trait MMSE and the pathological trait NFT. Table 6 provides a list of all clusters of genes in this experiment while considering both moderate AD samples along with incipient AD samples. Gene clusters were ranked based on average deviation of gene significance with respect to both traits. In the similar fashion, we carry out our analysis of identification of co-expressed genes clusters considering samples of severe AD along with samples of incipient AD. Table 7 provides a list of gene clusters ranked based on average deviation of gene significance with respect to both MMSE and NFT traits. The highest ranked gene cluster while considering severe AD and incipient AD samples is the SI-Black  $(M_1)$  containing 171 genes.

In order to identify the conservation of molecular function across disease severity, we find the common genes expressed across gene

Category	GO ID	Term	% comp	P-Value	Fold Enrichment	Bonferroni	FDR
BP	GO:0008380	RNA splicing	6.97	0.020137	3.77	0.99	26.0314
сс	GO:0043228	non-membrane-bounded organelle	27.90	8.26E-04	1.95	0.15	1.0357
	GO:0043232	intracellular non-membrane-bounded organelle	27.90	8.26E-04	1.95	0.15	1.0357
	GO:0031981	nuclear lumen	19.76	0.001064	2.38	0.19	1.3317
	GO:0005829	cytosol	18.60	0.001357	2.43	0.24	1.6952
	GO:0005730	nucleolus	12.79	0.001661	3.21	0.28	2.0721
	GO:0030529	ribonucleoprotein complex	10.46	0.002838	3.61	0.44	3.5161
	GO:0070013	intracellular organelle lumen	20.93	0.0036	2.05	0.52	4.4405
	GO:0043233	organelle lumen	20.93	0.004588	2.00	0.61	5.6268
	GO:0031974	membrane-enclosed lumen	20.93	0.005601	1.96	0.68	6.8299
	GO:0000267	cell fraction	12.79	0.033246	2.06	0.99	34.6775
	GO:0005624	membrane fraction	10.46	0.039223	2.26	0.99	39.5857
	GO:0005626	insoluble fraction	10.46	0.047145	2.18	0.99	45.5686
MF	GO:0003723	RNA binding	15.11	6.00E-04	3.17	0.12	0.7669
	GO:0000166	nucleotide binding	24.41	0.020421	1.63	0.99	23.2424

**Table 4**: Enriched GO categories of the 101 genes of CI-Purple ( $M_1$ ) using DAVID analysis.

Rank	Pathway name	Impact Factor	No. of Input Genes in Pathway	Rank	ank Pathway name F		No. of Input Genes in Pathway
1	Chronic myeloid leukemia	5.424	3	21	Pathogenic Escherichia coli infection	1.638	1
2	Ribosome	5.243	3	22	Endometrial cancer	1.585	1
3	Notch signaling pathway	4.008	2	23	Non-small cell lung cancer	1.551	1
4	Vibrio cholerae infection	3.552	2	24	Acute myeloid leukemia	1.503	1
5	Huntington's disease	3.228	3	25	Regulation of actin cytoskeleton	1.466	2
6	Long-term depression	3.168	2	26	Glioma	1.389	1
7	Gap junction	2.761	2	27	Epithelial cell signaling in Helicobacter pylori infection	1.35	1
8	GnRH signaling pathway	2.652	2	28	p53 signaling pathway	1.35	1
9	Parkinson's disease	2.41	2	29	Renal cell carcinoma	1.337	1
10	Ubiquitin mediated proteolysis	2.191	2	30	Melanoma	1.313	1
11	Tight junction	2.178	2	31	Long-term potentiation	1.313	1
12	RNA polymerase	2.153	1	32	VEGF signaling pathway	1.301	1
13	Thyroid cancer	2.12	1	33	Cardiac muscle contraction	1.278	1
14	Insulin signaling pathway	2.116	2	34	Fc epsilon RI signaling pathway	1.255	1
15	Wnt signaling pathway	1.969	2	35	Adherens junction	1.255	1
16	Alzheimer's disease	1.856	2	36	ErbB signaling pathway	1.143	1
17	Proteasome	1.777	1	37	7 Prostate cancer		1
18	Bladder cancer	1.777	1	38	38 MAPK signaling pathway		2
19	Pathways in cancer	1.748	3	39	Toll-like receptor signaling pathway	1.03	1
20	Shigellosis	1.638	1	40	Melanogenesis	1.022	1

**Table 5**: KEGG pathway analysis of 101 genes in the CI-Purple  $(M_{i})$  using the tool Pathway Express.

clusters (CI-Purple  $(M_1)$ , MI-Black  $(M_1)$ , and SI-Black  $(M_1)$ ) that have highest ranks.

To find the common functions retained between incipient AD and moderate AD samples, we consider those 57 genes that are common between gene clusters CI-Purple  $(M_i)$  and MI-Black  $(M_i)$ . We subject

these genes to the gene enrichment analysis using the tool DAVID EASE (Table 8). We observe two BP categories, nine CC categories, and three MF categories. The MF retained was the RNA binding (MF: 8 genes, p-value= $5.4 \times 10^{-3}$ ) and nucleotide binding (MF: 13 genes, p-value= $3.1 \times 10^{-2}$ ), when compared to the MF found in Table 4 of the



Gene Cluster (Rank)	MMSE significance	NFT significance	Avg. Dev	# genes
Black (M <sub>1</sub> )	0.1952707	0.6482817	0.2265055	177
Dark Magenta (M <sub>2</sub> )	0.1581371	0.5549828	0.19842285	63
Cyan (M <sub>3</sub> )	0.2103996	0.4298741	0.10973725	70
Purple (M <sub>4</sub> )	0.2061424	0.3619163	0.07788695	150
Dark green	0.3510179	0.2028728	0.07407255	191
Green-Yellow	0.1403241	0.282228	0.07095195	120
Dark Orange	0.2773515	0.160049	0.05865125	200
Orange	0.1869104	0.2888293	0.05095945	180
Orang red4	0.2074065	0.2781194	0.03535645	188
Grey	0.2160465	0.16033646	0.02785502	11
Mid-night blue	0.2286879	0.266835	0.01907355	64
Dark Red	0.1779462	0.215044	0.0185489	2160
Royal blue	0.3608635	0.333594	0.01363475	122
Tan	0.165247	0.1864312	0.0105921	102
Magenta	0.1692694	0.1901007	0.01041565	162
Salmon	0.393322	0.37553556	0.00889322	77
Light cyan	0.2969456	0.3096593	0.00635685	50
Pink	0.176833	0.1724148	0.0022091	163
blue	0.2299296	0.2284175	0.00075605	233

Table 6: Ranked gene clusters while taking into consideration samples of moderate AD along with samples of Incipient AD.

CI-Purple  $(M_i)$ . We then subject these common genes to identify the pathways associated with the common 57 genes using the tool pathway express. The resultant 41 pathways associated with these genes and their corresponding impact scores can be found in Table 9.

To identify those incipient co-expressed genes that are conserved through the severe samples; we extend our analysis to find the genes common to all three gene clusters CI-Purple  $(M_1)$ , MI-Black  $(M_1)$ , and SI-Black  $(M_1)$ . Table 10, contains the GO categories associated with

Gene Cluster (Rank)	MMSE significance	NFT significance	Avg. Dev	# genes
Black (M <sub>1</sub> )	0.2217576	0.4483017	0.11327205	171
Green yellow (M <sub>2</sub> )	0.3875915	0.218123	0.08473425	255
Purple (M <sub>3</sub> )	0.3251634	0.1886084	0.0682775	81
Turquoise (M <sub>4</sub> )	0.1960435	0.3198959	0.0619262	2518
Yellow	0.246421	0.1908917	0.02776465	202
Magenta	0.3314159	0.3810918	0.02483795	72
Brown	0.1498229	0.1884388	0.01930795	209
Green	0.3060369	0.3441088	0.01903595	188
Red	0.1859851	0.2000906	0.00705275	169
Grwy	0.2709881	0.2570546	0.00696675	5
Pink	0.2648702	0.2765768	0.0058533	613

Table 7: Ranked gene clusters by taking into consideration samples of severe AD along with samples of incipient AD.

Category	GO-ID	Term	Count	P-Value	Benjamini	FDR
BP	GO:0019941	modification-dependent protein catabolic process	5	0.0989	1	76.900
	GO:0043632	modification-dependent macromolecule catabolic process	5	0.0989	1	76.900
сс	GO:0043232	intracellular non-membrane-bounded organelle	16	5.74E-04	0.0729	0.6680
	GO:0043228	non-membrane-bounded organelle	16	5.74E-04	0.0729	0.6680
	GO:0005730	nucleolus	8	0.00164	0.1030	1.9065
	GO:0031981	nuclear lumen	10	0.00904	0.3294	10.064
	GO:0005829	cytosol	9	0.01769	0.4451	18.81
	GO:0070013	intracellular organelle lumen	10	0.03220	0.5785	31.764
	GO:0043233	organelle lumen	10	0.03676	0.5613	35.427
	GO:0030529	ribonucleoprotein complex	5	0.03998	0.5367	37.90
	GO:0031974	membrane-enclosed lumen	10	0.04100	0.4988	38.67
MF	GO:0000287	magnesium ion binding	5	0.04880	0.92334	45.13
	GO:0003723	RNA binding	8	0.00543	0.56798	6.328
	GO:0000166	nucleotide binding	13	0.03107	0.91204	31.52

Table 8: Enrichment analysis of genes common to gene clusters CI-Purple (M,) and MI-Black (M,) using DAVID EASE.

the 40 common genes using the gene enrichment analysis using the tool DAVID EASE. Once again we observe that the MF RNA binding (MF: 11 genes, p-value= $2.0 \times 10^{-2}$ ) and nucleotide binding (MF: 6 genes, p-value= $2.0 \times 10^{-2}$ ), were retained across through AD samples from incipient, moderate, and severe. We then subject these 40 genes to the tool pathway express to identify the associated pathways with the genes. Table 11 contains a list of the 38 pathways associated with genes.

### Discussion

The overarching question being addressed in this work is what are (if any) the gene sets in incipient AD samples that are co-expressed. Our objective is to identify these co-expressed genes sets as they could be potential disease markers that could be traced through their differential expression. We believe that these markers could to some extent be responsible for the phenotypic differences through the progression of AD.

The identification of AD markers in its incipient stages is a challenge considering the complexity of the disease and its close association with the normal aging process. In this work we have used gene co-expression network analysis (WGCNA) which is used to identify networks of co-expressed genes. We believed that functionally co-expressed genes can be modeled as a complex network. Subnetworks (clusters) within this gene network are believed to have a high level of connectivity determined by correlated expression profiles. We therefore subjected these gene clusters to gene ontology enrichment studies to list prominent functional associated with the genes.

Most of the genes in the incipient stage of AD are associated with RNA binding, nucleotide binding, RNA metabolism and processing, and cofactor binding. Nucleotide and RNA binding were the most prominent molecular functions associated with the co-expressed genes in the incipient stage.

Interestingly, much recent independent work has underscored the important role of up-regulated small non-coding RNAs (sncRNAs) called micro RNAs (miRNAs) in contributing to altered RNA signal processing, resulting in an altered transcriptome in AD brain [24,25].

The pathways associated with the co-expressed genes in the incipient stage of AD, are the chronic myeloid leukemia, ribosome, Notch signaling pathway, and the *Vibrio cholera* infection pathways. These four pathways were ranked the highest based on their impact scores.

Rank	Pathway Name	Impact Factor	#Input Genes in Pathway	Rank	Pathway Name	Impact Factor	#Input Genes in Pathway
1	Long-term depression	4.333	2	21	Melanoma	1.87	1
2	Ribosome	4.185	2	22	Long-term potentiation	1.87	1
3	Gap junction	3.902	2	23	VEGF signaling pathway	1.857	1
4	GnRH signaling pathway	3.785	2	24	Cardiac muscle contraction	1.832	1
5	Insulin signaling pathway	3.202	2	25	Chronic myeloid leukemia	1.819	1
6	Alzheimer's disease	2.911	2	26	Fc epsilon RI signaling pathway	1.807	1
7	Huntington's disease	2.793	2	27	Adherens junction	1.807	1
8	Thyroid cancer	2.717	1	28	ErbB signaling pathway	1.684	1
9	Regulation of actin cytoskeleton	2.463	2	29	Prostate cancer	1.664	1
10	Proteasome	2.362	1	30	Toll-like receptor signaling pathway	1.559	1
11	Bladder cancer	2.362	1	31	Melanogenesis	1.551	1
12	Shigellosis	2.216	1	32	T cell receptor signaling pathway	1.5	1
13	Pathogenic Escherichia coli infection	2.216	1	33	Parkinson's disease	1.444	1
14	Endometrial cancer	2.16	1	34	Leukocyte transendothelial migration	1.444	1
15	Non-small cell lung cancer	2.124	1	35	Natural killer cell mediated cytotoxicity	1.35	1
16	Acute myeloid leukemia	2.073	1	36	Ubiquitin mediated proteolysis	1.33	1
17	Vibrio cholerae infection	2.026	1	37	Tight junction	1.324	1
18	Glioma	1.951	1	38	Calcium signaling pathway	1.082	1
19	p53 signaling pathway	1.909	1	39	Focal adhesion	0.979	1
20	Renal cell carcinoma	1.896	1	40	MAPK signaling pathway	0.76	1
21	Melanoma	1.87	1	41	Pathways in cancer	0.616	1

**Table 9**: KEGG pathway analysis of 57 common genes between gene clusters CI-Purple ( $M_{\tau}$ ) and MI-Black ( $M_{\tau}$ ) using the tool Pathway Express.

Category	GO-ID	Term	Count	P-Value	Benjamini	FDR
сс	GO:0005829	cytosol	8	0.010332	0.6674	11.000
	GO:0043232	intracellular non-membrane-bounded organelle	11	0.011951	0.4712	12.620
	GO:0043228	non-membrane-bounded organelle	11	0.011951	0.4712	12.620
	GO:0005730	nucleolus	5	0.041082	0.7728	37.544
MF	GO:0000166	nucleotide binding	11	0.020211	0.9204	21.004
	GO:0003723	RNA binding	6	0.0205	0.7231	21.272

Table 10: Enrichment analysis of the genes common to gene clusters CI-Purple (M<sub>1</sub>), MI-Black (M<sub>1</sub>), and SI-Black (M<sub>1</sub>) using DAVID EASE.

Genes that are conserved through the progression of AD from incipient to moderate, include ACTB gene, the MAP2K2, NDUFV1, and GNA11 respectively. The pathways associated with gene ACTB include, regulation of atin cytoskeleton, shigellois, pathogenic Escherichia coli infection, *Vibrio cholera* infection, cardiac muscle contraction, aderens junction, leukocyte trans-endothelial migration, tight junction, and focal adhesion.

The gene NDUFV1, is associated with the following pathways: Alzheimer's disease, Huntington's disease, and Parkinson's disease.

The gene MAP2K2 is associated with a range of pathways that include, long term depression, gap junction, GnRH signaling pathway, insulin signaling pathway, thyroid cancer, regulation of actin cytoskeleton, bladder cancer, endometrial cancer, non-small cell lung cancer, acute myeloid cancer, glioma, renal cell carcinoma, melanoma, long-term potential, VEGF signaling pathway, chronic myeloid leukemia, Fc epsilon RI signaling pathway, ErbB signaling pathway, prostate cancer, toll-like receptor signaling pathway, melanogenesis, t-cell receptor signaling pathway, natural killer cell mediated cytotoxicity, MAPK signaling pathway, and pathways in cancer.

The most important is the gene GNA11 and its associated pathways relate to long term depression, gap junction, the calcium signaling pathway, and the GnRH signaling pathway (Figure 5).

Similarly, genes that are functionally significant and at the same time conserved throughout the progression of AD from incipient through severe include: RPL12, RPL13, FASN, GAPDH, CDC34, DCTN1, GNA11, ACTB, and MAP2K2. The corresponding pathways that gene RPL12 and RPL13 belong to Ribosome. Similarly, gene FASN belongs to the Insulin signaling pathway, gene GAPDH belongs to the Alzheimer's disease pathway, CDC34 belongs to the Ubiquitin mediated proteolysis pathway, and gene DCTN1 belongs to the Huntington's disease pathway. The gene GNA11 belongs to four pathways that

Rank	Pathway Name	Impact Factor	#Input Genes in Pathway	Rank	Pathway Name	Impact Factor	#Input Genes in Pathway
1	Long-term depression	5.035	2	20	Chronic myeloid leukemia	2.156	1
2	Gap junction	4.594	2	21	Fc epsilon RI signaling pathway	2.144	1
3	GnRH signaling pathway	4.475	2	22	Adherens junction	2.144	1
4	Insulin signaling pathway	3.874	2	23	Ribosome	2.096	1
5	Regulation of actin cytoskeleton	3.103	2	24	ErbB signaling pathway	2.017	1
6	Thyroid cancer	3.07	1	25	Prostate cancer	1.996	1
7	Bladder cancer	2.71	1	26	Toll-like receptor signaling pathway	1.888	1
8	Shigellosis	2.562	1	27	Melanogenesis	1.879	1
9	Pathogenic Escherichia coli infection	2.562	1	28	T cell receptor signaling pathway	1.826	1
10	Endometrial cancer	2.505	1	29	Leukocyte transendothelial migration	1.768	1
11	Non-small cell lung cancer	2.468	1	30	Natural killer cell mediated cytotoxicity	1.67	1
12	Acute myeloid leukemia	2.417	1	31	Ubiquitin mediated proteolysis	1.649	1
13	Vibrio cholerae infection	2.368	1	32	Tight junction	1.642	1
14	Glioma	2.292	1	33	Alzheimer's disease	1.464	1
15	Renal cell carcinoma	2.235	1	34	Huntington's disease	1.406	1
16	Melanoma	2.208	1	35	Calcium signaling pathway	1.386	1
17	Long-term potentiation	2.208	1	36	Focal adhesion	1.276	1
18	VEGF signaling pathway	2.195	1	37	MAPK signaling pathway	1.036	1
19	Cardiac muscle contraction	2.169	1	38	Pathways in cancer	0.874	1

Table 11: KEGG pathway analysis of 40 common genes between gene clusters CI-Purple (M<sub>1</sub>), MI-Black (M<sub>1</sub>), and SI-Black (M<sub>1</sub>) using the tool Pathway Express.



Citation: Pradeep C, Prerna D, Lukiw WJ (2013) An Exploratory Analysis of Conservation of Co-Expressed Genes across Alzheimer's disease Progression. J Comput Sci Syst Biol 6: 215-227. doi:10.4172/jcsb.1000119



include long-term depression, gap junction, GnRH signaling pathway, and calcium signaling pathway.

The gene MAP2K2 is associated with 25 known pathways that include: long-term depression, gap junction, GnRH signaling pathway, insulin signaling pathway, regulation of actin cytoskeleton, thyroid cancer, bladder cancer, endometrial cancer, non-small cell lung cancer, acute myeloid leukemia, glioma, renal cell carcinoma, melanoma, long term potentiation, VEGF signaling pathway, chronic myeloid leukemia, Fc epsilon RI signaling pathway, ErbB signaling pathway, prostate cancer, toll-like receptor signaling pathway, melanogenesis, t-cell receptor signaling pathway, and pathways in cancer.

Of all the associated genes and pathways, the genes GNA11 and MAP2K2 are prominently ranked in an ascending order as the disease progresses and are closely related to Alzheimer's disease.

#### Acknowledgement

This project was supported by grants from the National Center for Research Resources (NCRR) (P20RR016456), the National Institute of General Medical Sciences (NIGMS) (P20GM103424) from the National Institutes of Health (NIH) and LEQSF (2011-14)-RD-A-16. Research on Alzheimer's disease in the Lukiw laboratory was supported through Translational Research Initiative Grants from LSUHSC, Alzheimer Association Investigator-Initiated Research Grant (IIRG-09-131729), and NIA Grants (AG18031) and (AG038834). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. We would also like to extend our thanks to Kittu Chaithanya for his assistance through this analysis.

#### References

- Shen J, Kelleher RJ 3<sup>rd</sup> (2007) The presenilin hypothesis of Alzheimer's disease: evidence for a loss-of-function pathogenic mechanism. Proc Natl Acad Sci USA 104: 403-409.
- 2. Ballatore C, Lee VM, Trojanowski JQ (2007) Tau-mediated neurodegeneration

in Alzheimer's disease and related disorders. Nat Rev Neurosci 8: 663-672.

- Golde TE, Miller VM (2009) Proteinopathy-induced neuronal senescence: a hypothesis for brain failure in Alzheimer's and other neurogegenerative diseases. Alzheimers Res & Ther 1: 5.
- Smith IF, Green KN, LaFerla FM (2005) Calcium dysregulation in Alzheimer's disease: recent advances gained from genetically modified animals. Cell Calcium 38: 427-37.
- Dunckley T, Beach TG, Ramsey KE, Grover A, Mastroeni D, et al. (2006) Gene expression correlates of neurofibrillary tangles in Alzheimer's disease. Neurobiol Aging 27: 1359-1371.
- Blalock EM, Geddes JW, Chen KC, Porter NM, Markesbery WR, et al. (2004) Incipient Alzheimer's disease: Microarray correlation analyses reveal major transcriptional and tumor suppressor responses. Proc Natl Acad Sci US A 101: 2173-2178.
- Lee HK, Hsu AK, Sajdak J, Qin J, Pavlidis P (2004) Coexpression Analysis of Human Genes across many Microarray Data sets. Genemoe Res 14: 1085-1094.
- Horvath S, Dong J (2008) Geometric Interpretation of Gene Coexpression Network Analysis. PLoS Comput Biol 4: e1000117.
- Presson AP, Sobel EM, Papp JC, Suarez CJ, Whistler T, et al. (2008) Integrated weighted gene co-expression network analysis with an application to chronic fatigue syndrome. BMC Syst Biol 2: 95.
- Ray M, Zhang W (2010) Analysis of alzheimer's disease severity across brain regions by topological analysis of gene co-expression networks. BMC Syst Biol 4: 136.
- Stuart JM, Segal E, Koller D, Kim SK (2003) A gene-coexpression network for global discovery of conserved genetic modules. Science 302: 249-255.
- Ray M, Ruan J, Zhang W (2008) Variations in the transcriptome of Alzheimer's disease reveal molecular networks involved in cardiovascular diseases. Genome Biol 9: R148.
- 13. Zheng Q, Wang XJ (2008) GOEAST: a web-based software toolkit for gene

ontology enrichment analysis. Nucleic Acids Res W358-363.

- Huang D, Sherman B, Kempicki R (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4: 44-57.
- Tarca AL, Draghici S, Khatri P, Hassan SS, Mittal P, et al. (2009) A Novel Signaling Pathway Impact Analysis. Bioinformatics 25: 75-82.
- Gómez Ravetti M, Rosso OA, Berretta R, Moscato P (2010) Uncovering molecular biomarkers that correlate cognitive decline with the changes of hippocampus' gne expression profiles in alzheimer's disease. PLoS One 5: e10153.
- 17. Langfelder P, Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9: 559.
- Langfelder P, Lou R, Oldham MC, Horvath S (2011) Is my network module preserved and reproducible. PLoS Comput Biol 7: e1001057.
- 19. Saugstad JA (2010) Micrornas as effectors of brain function with roles in ischemia and injury, neuroprotection, and neurodegeneration. J Cereb Blood

Flow Metab 30: 1564-1576.

- 20. Zhang B, Horvath S (2005) A General Framework for Weighted Gene Coexpression Network Analysis. Stat Appl Genet Mol Biol 4: 17.
- Langfelder P, Zhang B, Horvath S (2008) Defining clusters from a hierarchical cluster tree: the Dynamic tree cut library for R. Bioinformatics 24: 719-720.
- 22. Ginos MA, Page GP, Michalowicz BS, Patel KJ, Volker SE, et al. (2004) Identification of a gene expression signature associated with recurrent disease is squamous cell carcinoma of the head and neck. Cancer Res 64: 55-63.
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, et al. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 25: 25-29.
- Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M,(2004) The KEGG resource for deciphering the genome. Nucleic Acids Res D277-280.
- 25. Lukiw WJ (2007) Micro-RNA speciation in fetal, adult and alzheimer's disease hippocampus. Neuroreport 18: 297-300.