Report Number: WUCSE-2008-6

2008-01-01

Transcriptome analysis of Alzheimer's disease identifies links to cardiovascular disease

Monika Ray, Jianhua Ruan, and Weixiong Zhang

2008-6

Transcriptome analysis of Alzheimer's disease identifies links to cardiovascular disease

Authors: Monika Ray, Jianhua Ruan, Weixiong Zhang

Corresponding Author: mray@cse.wustl.edu

Type of Report: Other

Transcriptome analysis of Alzheimer's disease identifies links to cardiovascular disease

Monika Ray[1*], Jianhua Ruan[2*], Weixiong Zhang[1,3]

[1]Washington University School of Engineering,

Dept. of Computer Science and Engineering, St. Louis, MO 63130

[2]University of Texas at San Antonio,

Dept. of Computer Science, San Antonio, TX 78249

[3]Washington University School of Medicine,

Dept. of Genetics, St. Louis, MO 63110

Correspondence Email: zhang@cse.wustl.edu

[*] Co-first authors

Abstract

Understanding the pathogenesis in the early stages of late-onset Alzheimer's disease (AD) can help in gaining important mechanistic insights into this devastating neurodegenerative disorder. Integration of multiple computational approaches to address the different levels of information embedded in microarray data, such as networks of co-expressed genes, functional annotation modules, and cis-regulatory elements shared by

co-expressed genes, leads to greater understanding of complex diseases such as AD. We use our recently developed methods for co-expression network analysis (CoExp) and genome-wide motif identification (WordSpy) along with functional annotation clustering on single cell expression data to analyse AD. CoExp automatically identified 6 clusters/modules, each of which represented a biological process perturbed in AD. Interestingly, AD related genes like APOE, A2M, PON2, MAP4 and cardiovascular diseases (CVD) associated genes such as COMT, CBS, WNK1 all congregated in one of the six modules. This module that contained 18 disease associated (cardiovascular, neurodegeneration, diabetes, stroke) genes had the maximum number of hub genes. Some of the disease related genes were also hub genes while many of them were directly connected to one or more hub genes. Further investigation of this disease associated module unveiled significant cis-regulatory elements that were significantly similar to the binding sites of transcription factors involved in AD and CVD. Our results showcase extensive links between genes associated with AD and CVD at the co-expression and co-regulation levels and provide strong supporting evidence to the hypotheses linking CVD and AD.

1 Introduction

Late-onset Alzheimer's disease (AD) is a complex progressive neurodegenerative disorder of the brain and is the commonest form of dementia. A systems biology approach is an efficient way to analyse complex diseases, such as AD, that are polygenic and have multiple, interacting genes contributing to the spectrum of variation.

We perform a transcriptome-based analysis of AD by combining our newly developed co-expression network (CoExp) and genome-wide motif identification (WordSpy) methods to study

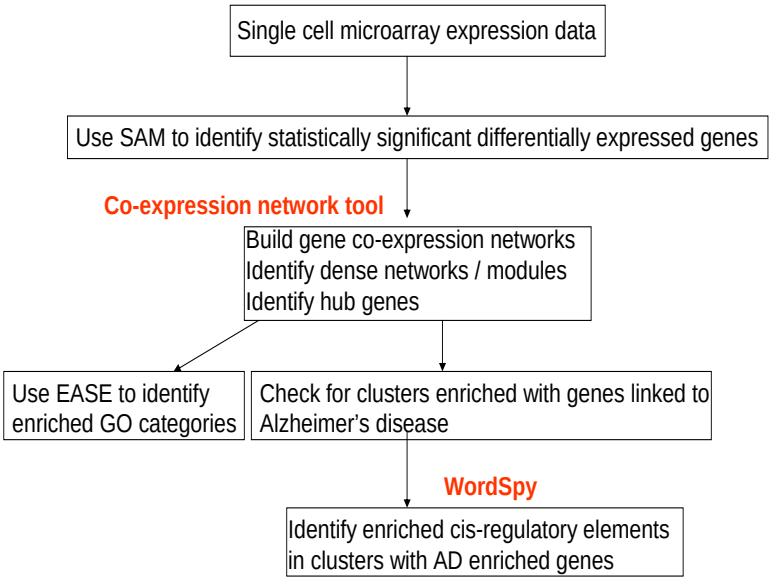

Figure 1: Sequence of steps taken to analyse incipient Alzheimer's disease (AD) from single cell expression data. We apply co-expression network analysis and WordSpy (motif finding) in an integrated manner to study AD and reveal connections to other conditions such as cardiovascular diseases and diabetes.

early response genes in AD(Ruan and Zhang, 2006b, 2008; Wang and Zhang, 2006). In the first stage, CoExp is used to construct modules of tightly correlated genes (i.e. high similarity in their expression profiles). In the next stage, WordSpy identifies regulatory cis-elements (motifs), which are then used to group genes within a module based on the motifs they share. The analysis follows the procedure shown in Figure 1.

The present work unveiled 1663 genes that are differentially expressed in AD. CoExp was applied to these genes resulting in 6 modules of co-expressed genes, each module representing key biological processes perturbed in AD. Within the 6 modules we identified 107 highly connected ('hub') genes, many of which play important roles in AD. Functional annotation clustering based on association to human diseases resulted in the identification of 18 disease related (cardiovascu-

lar diseases, AD/neurodegenerative diseases, stroke and diabetes) transcripts aggregating in one module (referred to as the disease associated module). Some of these 18 genes were also hub genes, while many of them were directly connected to one or more hub genes. Further analysis of the disease associated module using WordSpy[1], resulted in several cis-regulatory elements that matched binding sites of transcription factors involved in diseases that are known to co-occur with AD. The final result is a set of co-expressed and co-regulated modules describing the higher level characteristics linking AD and cardiovascular diseases.

Our work is significantly different from that by Miller et al.(Miller et al., 2008) as we use a different co-expression network building method (CoExp) to generate modules of co-expressed genes and then identify cis-regulatory motifs in a module. CoExp is a spectral algorithm that was designed to optimise a modularity function and automatically identify the appropriate number of modules(Ruan and Zhang, 2006b, 2008). The cis-regulatory elements discovered in the promoter regions of disease related genes provide further insights into the possible transcriptional regulation of the genes involved in AD and their connection to other diseases, such as cardiovascular diseases, stroke and diabetes. Moreover, the single cell dataset, which is less noisy compared to mixed cell microarray data, is more recent(Dunckley et al., 2006) compared to those analysed by Miller et al. Most importantly, unlike multiple studies comparing AD and ageing(Ricciarelli et al., 2004; Miller et al., 2008; Pereira et al., 2007), to the best of our knowledge, our study is the first that has identified links between cardiovascular diseases, AD/neurodegenerative diseases and diabetes using a transcriptome-based systems biology approach. Lastly, the AD expression data that we analyse are from the entorhinal cortex, a region of the brain known to be the germinal site of AD. Despite the differences between our study and that by Miller et al., we have established interesting

[1]CoExp and WordSpy can be retrieved from http://www.cse.wustl.edu/ ∼zhang/

links between the two studies, thereby highlighting the commonalities between AD, ageing, and cardiovascular diseases. We believe that analyses such as ours and that by Miller et al. are the pieces of a puzzle that will result in a more comprehensive understanding of complex diseases such as Alzheimer's and its link to other conditions/diseases.

2 Data and Methods

2.1 Data

Pathologically AD is characterised by the presence of neurofibrillary tangles (NFT) in the neurons of the entorhinal cortex and hippocampus. Dunckley et al. dataset consists of 13 normal controls (Braak stages 0–II; average age: 80.1 years) and 20 AD affected (Braak stages III–IV; average age: 84.7 years) samples obtained by laser capture microdissection (LCM) from the entorhinal cortex(Dunckley et al., 2006). Braak stages III–IV is considered 'incipient' AD(Rossler et al., 2002; Braak and Braak, 1991). In this dataset, 1000 neurons were collected from each of the 33 samples via LCM.

Data were normalised using gcRMA(Irizarry et al., 2006). Probesets were mapped to genes using DAVID(Dennis et al., 2003). Probesets that did not map to any gene name and those matching to hypothetical proteins with no known functions, at the time of writing this manuscript, were removed. When multiple probesets mapped to the same gene, only the probeset with the highest mean was selected. Differentially expressed genes were selected using significance analysis of microarrays (SAM)(Tusher et al., 2001).

2.2 Construction of co-expression networks

We use a network-based approach to identify modular structures/clusters embedded in microarray gene expression data. The co-expression network (CoExp) method constructs co-expression networks from microarray data and then uses a spectral based clustering method to identify subgraphs within the network(Ruan and Zhang, 2006b, 2008). Each node in the network is a gene and edges represent expression similarities between genes. The idea is that genes involved in the same functional pathway are directly connected to each other or linked via short paths. After the network is created, the nodes are clustered into different dense subgraphs. Most clustering algorithms require the user to specify the number of clusters/modules. However, CoExp uses a spectral based clustering algorithm that optimises the modular function proposed by Newman and Girvan(Newman and Girvan, 2004) to automatically determine the appropriate number of modules(Ruan and Zhang, 2006b, 2008). The CoExp is easily scalable to large networks. Further evidence of its robustness can be found in (Ruan and Zhang, 2008, 2006a).

EASE (http://niaid.abcc.ncifcrf.gov/home.jsp) was used to identify overrepresented biological processes in each module as well as perform functional annotation clustering based on association to human diseases.

2.3 Identification of regulatory cis-elements

The interaction of transcription factors (TFs) and cis-acting DNA elements determines the gene activity under various environmental conditions. Identifying functional TF binding sites, however, is not trivial, since TF binding sites are usually short and degenerate, and are often located several hundred to thousand bases upstream to the translational starting sites. Here we combine several

data sets and a whole-genome analysis method to discover short DNA sequence motifs that are statistically enriched in the promoters of genes in the same co-expression module and are associated with gene co-expression.

Briefly, we first download the promoter sequences for human ORFs from the DBTSS database (Wakaguri et al., 2008). Each promoter includes 1000bp upstream and 200bp downstream sequences relative to the transcription starting site, defined from full length cDNA data. From this data set we extract n sets of promoter sequences (referred to as experimental sets), where n is the number of co-expression modules. The i-th experimental set contains the promoter sequences of genes in the i-th co-expression module. The complete set of human gene promoters is used as the background set. We then run WordSpy, a steganalysis-based genome-wide motif-finding method(Wang and Zhang, 2006), on each experimental set in order to discover statistically significant k-mers (motifs) (for $k = 6, 7, 8, 9, 10$) according to a generative model of the promoter sequences.

Each k-mer identified by WordSpy is then subject to two filtering steps. In the first filtering step, we select motifs that are specifically enriched in the experimental set. We count the number of instances that a k-mer appears in the experimental set (denoted by x) and in the background set (denoted by b). Then we compute the probability that we would expect by chance at least the same number of occurrences in the experimental set given the number of occurrences in the background set. This probability is computed using the cumulative hyper-geometric distribution as follows -

$$P(x, b, N_i, N) = \sum_{k=x}^{min\{x,b\}} \frac{\binom{N_i}{k}\binom{N-N_i}{b-k}}{\binom{N}{x}} \tag{1}$$

where N_i and N are the sizes of the i-th experimental set and the background set, respectively. We filter out the k-mers that have a p-value ≥ 0.01.

The second filter is used to select motifs that are associated with strong and significant co-expression patterns. For each motif remaining after the first filtering phase, we obtain a set of genes ('target set') in which each gene in this set contains the motif in its promoter region. We compute the average pair-wise Pearson correlation coefficients, denoted by *pcc*, from the expression profiles of the genes in the target set. Furthermore, we randomly sample 100 control sets of genes from the background set that have the same size (i.e. number of genes) as the target set, and compute the *pcc* of each control set. The mean and standard deviation (denoted by *mpcc* and *spcc*, respectively) of the *pcc* values for the control sets are then used to compute the Z-score of the *pcc* value for the target set as follows -

$$Zscore = \frac{(pcc - mpcc)}{spcc} \tag{2}$$

A motif is retained only if its *pcc* > 0.4, and its Z-score > 2.

Finally, the motifs that have passed both filters are compared to the known TFBS in the JASPAR database(Sandelin et al., 2004). We pre-filter the TFBSs in the databse that have information content ≤ 6 bits, since these TFBSs are short and have high degeneracy and, hence, may match to some known motifs by chance. Then we compute the best un-gapped alignment between the motifs (n-mers) and the known binding sites (position specific weight matrices) using a metric called the *information score*, which is the same metric used in the Matlnspector program(Quandt et al., 1995) included in the TRANSFAC suite. We consider a motif match to a known TFBS if the information score is ≥ 0.8.

3 Results and Discussion

SAM identified 1663 differentially expressed (DE) genes between AD samples and controls at a false discovery rate (FDR) of 0.11%(Tusher et al., 2001). All the enriched biological processes of the transcripts is shown in Table S1 in the supplemental data files. Many of the processes known to be affected in Alzheimer's are enriched in the list of 1663 transcripts.

3.1 Modular organisation of significant genes via co-expression networks

The co-expression network method (CoExp) was applied to the set of 1663 genes and resulted in 6 modules(Ruan and Zhang, 2006b, 2008). Figure 2 shows the resulting adjacency matrix and Figure 3 shows the co-expression network.

The two disconnected groups of modules in Figure 3 represent 2 groups of anti-correlated expression patterns. Transcripts in modules 3,4,5 and 6 are downregulated and those in modules 1 and 2 are upregulated. The group on the left in Figure 3 (modules 1 and 2) contains many transcripts involved in cell differentiation, neuron development, immune response, stress response, etc., while the group on the right contains genes involved in negative regulation of metabolism, protein transport, sodium ion transport, etc. Table 1 shows the top enriched GO biological processes ($p < 0.05$) in all 6 modules.

As can be noted from Table 1, many processes linked to Alzheimer's, such as immune response, inflammatory response, cell development and differentiation (due to a large number of cancer related genes), etc. are upregulated in incipient AD(Norris et al., 2005; Y et al., 2001). Processes related to actin are downregulated in AD(Kojima and Shirao, 2007). Table 2 shows the most significant KEGG pathways represented by the genes in each module. Although, there was no

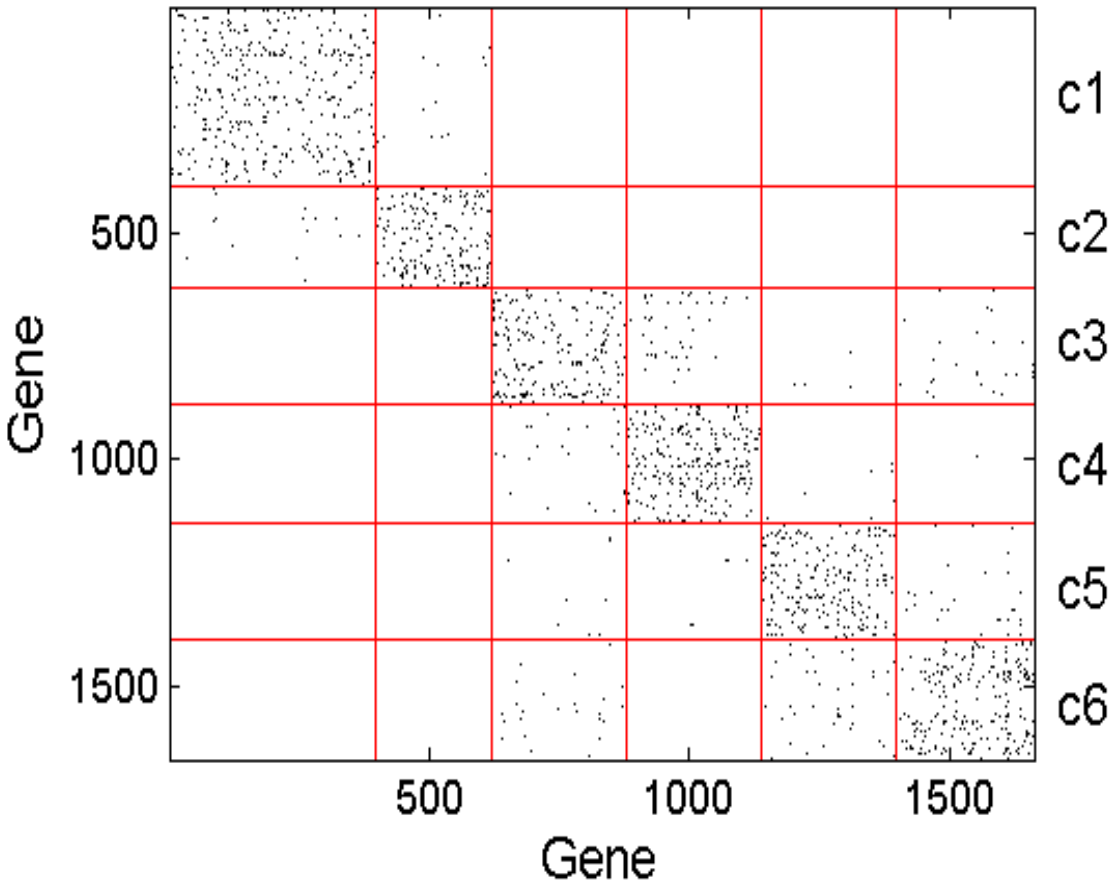

Figure 2: The adjacency matrix representation of the coexpression network. The graphical representation of this matrix is in Figure 3. Modules are labelled c1, c2, c3, c4, c5 and c6. Modules c1 and c2 refer to the group on the left in Figure 3 and the modules c3, c4, c5 and c6 refer to the one on the right. The dots refer to the intra- and inter-module edges between the genes.

Table 1: Top enriched GO biological processes in each module ($p < 0.05$)

Module	**Activity**	**Ease score**
Module 1	Protein biosynthesis	7.14E-06
	Cell development	2.37E-05
	Cell differentiation	4.88E-05
	Macromolecule biosynthesis	8.56E-05
	Cellular nerve ensheathment	1.11E-04
	Neuron development	2.22E-04
	Regulation of action potential	4.37E-04
Module 2	Response to other organism	0.004
	Immune response	0.014
	Defense response	0.020
	Response to stress	0.029
	Protein kinase cascade	0.030
	Integrin-mediated signalling pathway	0.030
	Myeloid cell differentiation	0.040
	JAK-STAT cascade	0.042
Module 3	Homophilic cell adhesion	2.58E-11
	Cell-cell adhesion	2.74E-09
	Nervous system development	3.44E-09
	Ion transport	0.007
	Gamma-aminobutyric acid signalling pathway	0.009
	Secretory pathway	0.019
	Small gtpase mediated signal transduction	0.028
	Sodium ion transport	0.036
Module 4	Cellular physiological process	6.91E-05
	Transcription from RNA polymerase II promoter	0.008
	Protein transport	0.014
	Post-chaperonin tubulin folding pathway	0.019
	Ubiquitin cycle	0.037
Module 5	Negative regulation of metabolism	0.011
	Actin filament depolymerisation	0.025
	Barbed-end actin filament capping	0.025
	Negative regulation of actin filament depolymerisation	0.025
	Negative regulation of protein metabolism	0.025
Module 6	Protein transport	0.008
	Cell organisation and biogenesis	0.011
	Membrane fusion	0.028
	RNA processing	0.029
	RNA splicing	0.042

Table 2: Over-represented KEGG pathways in each module ($p < 0.05$).

Module	**KEGG pathway**	**Ease score**
Module 1	Ribosome	8.16E-07
	Translation	3.41E-14
Module 2	Phospholipid degradation	0.013
Module 3	Signal transduction	0.002
	Phosphatidylinositol signalling system	0.005
Module 4	Neuron development	2.22E-04
Module 6	Nucleotide metabolism	0.036

over-represented KEGG pathway in module 5, several genes involved in the negative regulation of metabolism, actin filament depolymerisation, glucose metabolism, lipid biosynthesis were present. Modules 2,3,4,5 and 6 represent processes previously associated to AD in multiple studies(Norris et al., 2005; Y et al., 2001; Kojima and Shirao, 2007). Module 3,4,5 and 6 contain genes that have decreased expression levels. In particular, module 5 contains processes related to glucose metabolism. Recent work has shown decreased expression of energy metabolism genes(Liang et al., 2008). Our results further confirm this observation. Based on the results obtained thus far, each cluster/module can be a representative of some biological processes - module 1 represents protein synthesis, module 2 is linked to phospholipid degradation, module 3 is associated with signalling systems, module 4 represents neuron development and modules 5 and 6 are associated with metabolism.

The modular organisation of genes led to the following investigative steps - (a) the identification of module(s) associated with human diseases, (b) the identification of hub / highly connected genes within the modules, (c) the examination of the expression level of brain derived neurotrophic factor in the AD subjects, and (d) the identification of cis-regulatory elements from the promoters of

genes.

This prompted an in-depth examination of module 1.

3.1.1 Module associated with cardiovascular diseases and diabetes

When EASE (http://niaid.abcc.ncifcrf.gov/home.jsp) was used to perform functional annotation clustering based on the genes' association to human disorders/diseases, module 1 was the only module that had 18 disease associated genes (see Table 3). The connection to human conditions and diseases was made by EASE using the Genetic Association Database (geneticassociationdb.nih.gov). Modules 2-6 did not have a significant enrichment for any human disease.

Table 3: Functional annotation clustering of genes in module 1 based on association to human conditions/diseases

Disease/condition	**Genes**
Neurodegeneration	VWF, A2M, APOE, FTL, PON2, COMT, MAP4, TF, SERPINA3, ATP1A2, AGT
Myocardial infarction	A2M, APOE, PON2, SERPINA3
Alzheimer's disease	A2M, APOE, SERPINA3, PON2
Cardiovascular	VWF, A2M, APOE, PON2, COMT, WNK1, CBS, SERPINA3, TIMP1
Coronary artery disease	APOE, PON2, COMT, SERPINA3
Type 2 Diabetes	VWF, A2M, APOE, PCBD2, HLA-DQB1(HLA-DQB2), TIMP3, SLC2A1, AGT

These results provide a new set of evidence supporting the hypothesis that there may be a strong association between cardiovascular disease (CVD) and the incidence of Alzheimer's disease(STAMPFER, 2006; Rosendorff et al., 2007; STEWART, 1998). There also has been a growing body of evidence for a link between AD and diabetes(Janson et al., 2004; MacKnight et al., 2002; Craft and Watson, 2004), with many research groups and news articles reporting that AD may be another form of diabetes. While there are many shared transcripts in Table 3 among the

different conditions, there are a few that are unique to the disease/condition, such as kinase deficient protein (WNK1), timp metallopeptidase inhibitor 1 (TIMP1) and cystathionine-beta-synthase (CBS) which are specific to CVD, and pterin-4 alpha-carbinolamine dehydratase/dimerization cofactor of hepatocyte nuclear factor 1 alpha (tcf1) 2 (PCBD2), timp metallopeptidase inhibitor 3 (TIMP3), solute carrier family 2 (facilitated glucose transporter), member 1 (SLC2A1) and major histocompatibility complex, class ii, dq beta 1 (HLA-DQB1) being specific to diabetes. On the other hand, von willebrand factor (VWF), alpha-2-macroglobulin (A2M), apolipoprotein e (APOE), paraoxonase 2 (PON2), and serpin peptidase inhibitor, clade a (alpha-1 antiproteinase, antitrypsin), member 3 (SERPINA3) are common to most of the conditions. This indicates that genes that are specific to CVD and diabetes, conditions commonly associated with AD, were identified to be in the same module as genes related to neurodegenerative disease, including AD. Since the motivation behind co-expression network analysis is to identify possibly co-regulated genes, it is possible that these genes are co-regulated. Since genes that are common as well as those that are specific to the conditions in Table 3 are being co-regulated, it may be the reason for the clustering of these conditions in epidemiological studies. Furthermore, as there are many transcripts common to these diseases/conditions, it is plausible that similar/common biochemical pathways are active in these seemingly different conditions. If this is the case, then common pathogenetic mechanisms for AD and CVD can suggest a causal link between CVD and AD(Rosendorff et al., 2007; STEWART, 1998), a hypothesis that is still controversial and under a lot of debate.

Transcripts in the modules are linked to each other based on their expression similarity. 'Hub genes' are highly-connected nodes/transcripts in the network and are likely to play important roles in the biological processes. Hub genes tend to be conserved across species and hence, make excellent candidates for disease association studies in humans(Casci, 2006).

We defined hub genes to be those genes that have 40 or more links/connections. This resulted in 107 hub genes. The complete list of hub genes, their module locations, and the number of links is in Supplemental Table S2. The hub genes included general transcription factor iiic, polypeptide 1, alpha 220kda (GTF3C1) which is involved in RNA polymerase III-mediated transcription, microtubule-associated protein 4 (MAP4) which promotes microtubule stability and affects cell growth(Nguyen et al., 1998), and proprotein convertase subtilisin/kexin type 2 (PC2) which is responsible for the processing of neuropeptide precursors. Some of these hub genes - PC2, paraoxonase 2 (PON2) and peroxiredoxin 6 (PRDX6) - have been implicated in late-onset AD (LOAD)(Krapfenbauer et al., 2003; Shi et al., 2004; Winsky-Sommerer et al., 2003).

Since module 1 has the disease associated genes, the identification of hub genes in this module may provide new information regarding AD, CVD and diabetes. We identified 22 hub genes with the number of links ranging from 42 to 63 in module 1 (For the complete list of 22 hub genes, see Table S2). The total number of hub genes in each module along with the minimum and maximum number of links is shown in Table 4.

Table 4: Number of hub genes and their range of connections/links in each module

Module	**No. hubs**	**Range of links**
Module 1	22	42 – 63
Module 2	17	41 – 56
Module 3	15	40 – 68
Module 4	14	40 – 65
Module 5	20	40 – 73
Module 6	19	40 – 81

As shown in Table 4, module 1 had the maximum number of hub genes. The transcript with the largest number of links in module 1 is MAP4 with 63 connections. Table 5 shows the number of links for the 18 disease associated genes.

Table 5: Number of links of the 18 disease associated genes from module 1 and the number of connections they have with other hub genes

Gene	**No. links**	**No. hub genes it is connected to**
VWF	16	2
A2M	17	3
APOE	18	3
FTL	18	3
PON2	51	8
COMT	17	0
MAP4	63	1
TF	16	3
SERPINA3	18	3
ATP1A2	45	6
AGT	27	5
TIMP1	14	2
WNK1	17	1
CBS	16	3
PCBD2	16	0
HLA-DQB1	15	2
SLC2A1	14	1
TIMP3	14	0

It can be seen from Table 5, that PON2, MAP4 and atpase Na+/K+ transporting, alpha 2 (+) polypeptide (ATP1A2) are hub genes. The overexpression of MAP4 results in the inhibition of organelle motility and trafficking(Bulinski et al., 1997) and can also lead to changes in cell growth(Nguyen et al., 1998). ATP1A2 is a subunit of an integral membrane protein which is re-

sponsible for establishing and maintaining the electrochemical gradients of Na and K ions across the plasma membrane(Dennis et al., 2003). These gradients are essential for osmoregulation, for sodium-coupled transport of a variety of molecules, and for electrical excitability of nerve and muscle(Dennis et al., 2003). While the downregulation of ATP1A2 has been linked to migraine related conditions(De Fusco et al., 2003), the effects of its upregulation has not been documented. PON2 has been implicated in AD(Shi et al., 2004) and cardiovascular diseases (Table 3).

MAP4 is directly linked (linkage implies similarity in gene expression profile) to other disease/condition associated genes such as VWF and WNK1. Increased expression of semaphorin 3b (SEMA3B) (semaphorin pathway) inhibits axonal elongation(Blalock et al., 2004), which has been implicated in AD(Blalock et al., 2004). MAP4 is also connected to SEMA3B. Many of the 18 disease associated genes are linked to one or more hub genes (see Table 5). Although, not all the disease associated genes are hub genes themselves, most of them are directly linked to one or more hub genes, which implies that they may play a role via hub genes in the biological processes represented by the module.

3.1.2 Decreased levels of brain-derived neurotrophic factor

Brain-derived neurotrophic factor (BDNF) is well known for its trophic functions and has been implicated in synaptic modulation, and the induction of long-term potentiation (LTP)(Yamada et al., 2002; Tyler et al., 2002). Decreased levels of BDNF has been linked to Alzheimer's and depression(Tsai, 2003; Laske et al., 2006; Karege et al., 2002). Recently, low levels of BDNF has also been associated with diabetes(Krabbe et al., 2007).

BDNF goes through post-translational modification i.e. converted into mature BDNF by plasminogen (PLG) (http://www.genecards.org). The neurotrophic tyrosine kinase, receptor, type 2

(NTRK2/TrkB) is a receptor for BDNF(Haapasalo et al., 2002).

BDNF was not present in our list of 1663 significant genes. However, TrkB and serpin peptidase inhibitor, clade e (nexin, plasminogen activator inhibitor type 1), member 2 (SERPINE2) were present in the set of 1663 genes and located in module 1. Plasminogen activator inhibitor type 1 (PAI-1) proteins inhibit plasminogen (PLG) activators(Huber et al., 2001). Therefore, if the level of PAI-1 is high in the AD affected samples, plasminogen activators are being inhibited, resulting in decreased levels of mature BDNF. Interestingly, the expression levels of TrkB and PAI-1 were elevated in the AD samples. However, TrkB is downregulated following the binding of BDNF(Sommerfeld et al., 2000). Therefore, due to an increased level of PAI-1, mature BDNF could not be produced, which in turn could not bind to TrkB. Hence, from this analysis it can be concluded that high levels of TrkB and PAI-1 implies decreased levels of BDNF, which is very critical for the survival of neuronal populations. This probably leads to neuronal death in this cohort of AD affected subjects.

In order to verify our conclusion regarding the expression level of BDNF in the AD patients in our dataset, we examined the expression level of BDNF in the controls and AD affected samples. We found BDNF to be decreased by 1.07 in the AD affected samples. BDNF was not selected to be a significant gene probably because it had a small difference in the expression between controls and affected samples. Microarrays are not sensitive enough to detect genes with low expression levels, especially when the difference in expression is small (which can be expected in subjects with incipient AD)(Bunney et al., 2003; Pan et al., 2006; Yue et al., 2001; Canales et al., 2006). The fact that the selected significant genes, such as TrkB and SERPINE2, could lead to the correct conclusion regarding the level of BDNF expression in AD affected samples, highlights the merits of this kind of analysis of the transcriptome when handling genes with low expression levels.

Although modules 1 and 2 have upregulated genes, genes associated with BDNF are located only in module 1. This further emphasises the importance of module 1.

3.1.3 Comparison to the study by Miller et al. on ageing and AD

Miller et al. identified 558 transcripts that were common to AD and ageing(Miller et al., 2008). We found more overlapping genes between our study and their's than expected by chance ($p = 3.3$ X 10^{-10}). 94 genes overlapped between 1663 significant genes from our study and 558 genes identified by Miller et al.. Of these 94 genes, 48 were present in module 1 (greater than expected by chance $p = 9.2$ X 10^{-10}), while the overlap between 558 AD-ageing genes and genes in modules 2,3,4,5 and 6 ranged between 2 - 15. This indicates that module 1 contains the majority of genes that have been linked to ageing and AD. Of the 48 genes that overlapped between 558 AD-ageing common genes and genes in module 1, WNK1 and MAP4 were present. MAP4 is associated with neurodegeneration while WNK1 is linked to cardiovascular diseases.

Furthermore, 9 genes (DAAM2, EPM2AIP1, GFAP, GORASP2, MAP4, NFKBIA, PRDX6, TSC22D4 and UBE2D2) overlapped between 558 AD-ageing genes and the 107 hub genes identified in our study, 5 of which resided in module 1. These results further emphasise the significance of module 1. From these results, it can be concluded that module 1 is an important link in the chain of pathophysiological characteristics connecting AD, CVD and ageing. It represents common biochemical pathways that may be affected in all these conditions.

3.2 Cis-regulatory elements and co-regulated genes

Cis-regulatory elements/motifs are regulatory elements in the promoter region of genes to which transcription factors (TFs) bind, and regulate transcription. If a group of genes share the same

cis-regulatory motif, then the TF that binds to the motif may regulate this group of genes. Co-expressed modules represent genes that may be co-expressed in the cell and be a part of the same biochemical pathways. Throughout the results in section 3.1 of our analysis, we observe that the genes contained in module 1 is of high importance. Since module 1 is the most important module identified in this analysis, we attempted to identify the cis-regulatory elements/motifs that may be enriched in the upstream promoter sequences of the genes in module 1. The group of genes in module 1 that share a motif will be a set that is co-expressed and co-regulated.

The complete set of cis-regulatory elements enriched in module 1 is in Supplemental Table S3. A total of 89 motifs were specifically enriched in module 1 with a p-value < 0.001, and their target genes were co-expressed with an average correlation coefficient > 0.4 and Z-score > 2 (see Data and Methods). Of the 89 motifs, 36 matched to 26 known transcription factor binding sites (TFBS) in JASPAR (http://jaspar.genereg.net/) with a matching score ≥ 0.8 (see Table 6). Table 6 shows the number of genes within module 1 whose promoter region contains a motif that matched to the TFBS of a known TF.

Transcription factors such as growth factor independent (Gfi), peroxiredoxin 2 (Prx2/PRDX2), SP1, CAAT-enhancer binding protein C/EBP), RelA(p65), runt box 1 (Runx1), ELK-1, upstream stimulatory factor 1 (USF1), Rel, TATA box binding protein (TBP) have been implicated in neurodegenerative diseases (such as AD, Parkinson's, and Schizophrenia)(Tsuda et al., 2005; Qu et al., 2007; Fang et al., 2007; Santpere et al., 2006; Christensen et al., 2004; Li et al., 2004; Perez-Capote et al., 2006; Barkett and Gilmore, 1999; Tomita et al., 2000; Kimura et al., 2007; Pastorcic and Das, 2003; Tong et al., 2004; Salero et al., 2003; Reid et al., 2004), diabetes(Ng et al., 2005), stroke and cardiovascular diseases(Choquette et al., 2007; Komulainen et al., 2006). 139 genes in module 1 contained motifs that matched the TFBS of the known TFs associated with these diseases.

Table 6: The 26 known TFs and the number of target genes in module 1 that have a motif in their promoters that match to the binding sites of the known TF

26 Transcription factors	No. target genes
ABI4	9
Arnt-Ahr	93
ARR10	6
Broad-complex_3	10
cEBP	20
Gfi	8
HAND1-TCF3	279
Mycn	11
Myf	8
Prx2/PRDX2	17
RELA, REL	10
RUNX1	4
Snail	49
SP1	47
TBP	6
E74A	16
ELK1	16
SPIB	16
Hunchback	6
MAX	11
USF1	11
ZNF42_5–13	27
NFIL3	5
Agamous	8
GAMYB	6

Arnt-Ahr dimer transcription factor activates genes crucial in the response to hypoxia and hypoglycaemia(Maltepe et al., 1997; Erbel et al., 2003). Hypoglycaemia and hypoxia have been known to play pathophysiological roles in the complications of diabetes and AD(Catrina et al.,

2004; Shi et al., 1997; Peers et al., 2007; Sun et al., 2006). It is well known that hypoxia has major effects on the cardiovascular system(Germack et al., 2002). Therefore, in light of such knowledge, it is no surprise that a large number of genes have cis-regulatory motifs that match the binding site of the Arnt-Ahr TF.

Hand1-TCF3 and TAL1-TCF3 are components of the basic-helix-loop-helix (bHLH) complexes. BHLH transcription factors are important in development (Yelon et al., 2000; Firulli et al., 2003). An extremely high number of genes were mapped to Hand1-TCF3 since cell development and differentiation is upregulated in AD (Norris et al., 2005; Y et al., 2001).

In summary, the fact that TFs which are active in other human diseases/disorders have their binding motifs enriched in the set of significant genes associated with AD adds significance to the hypothesis that many common biochemical pathways are affected in AD and CVD.

4 Conclusion

In this study, we presented an integrative systems biology approach to study a complex disease such as Alzheimer's disease. Along with identifying modules / clusters that illuminate higher-order properties of the transcriptome, we identified a module that contained many genes known to play prominent roles in cardiovascular diseases and AD. We identified several cis-regulatory elements, some of which mapped to the binding sites of known TFs involved in neurodegenerative and cardiovascular diseases as well as diabetes and stroke. Furthermore, since microarrays are not sensitive to genes with very slight differences in expression from controls, we illustrated how *other* genes can be used to deduce the expression difference of such genes. This is especially critical while comparing groups that are very similar to each other.

The link between cardiovascular diseases, diabetes and AD is a topic of growing interest. The presence of genes and cis-regulatory elements related to cardiovascular diseases and AD in a single module, provides strong evidence to the hypotheses connecting these two conditions. Interestingly, this module also contained the maximum number of genes (and hub genes) related to ageing. Our results support the notion that diseases in which the same set of biochemical pathways are affected may tend to co-occur with each other. This may be the reason why cardiovascular diseases and/or diabetes co-occur with AD. A comprehensive analysis incorporating AD and CVD/diabates patients along with information about their disease progression will lead to a more powerful analysis. Such a study will shed more light into the pathophysiology of AD and associated diseases.

Acknowledgement

The research was supported in part by NSF grant IIS-0535257 and a grant from the Alzheimer's Association. JR was supported by University of Texas, San Antonio new faculty startup grant and a faculty research award. The authors would like to thank Jeremy Miller at the Interdepartmental Program for Neuroscience and Centre for Neurobehavioural Genetics, University of California, Los Angeles, CA for his assistance in obtaining data from his AD-ageing paper.

References

Barkett, M. and Gilmore, T. D., 1999. Control of apoptosis by rel/nf-kappab transcription factors. *Oncogene*, **18(49)**:6910–6924.

Blalock, E. M., Geddes, J. W., Chen, K. C., Porter, N. M., Markesbery, W. R., and Landfield, P. W., 2004. Incipient alzheimer's disease: Microarray correlation analyses reveal major transcriptional and tumor suppressor responses. *Proc Natl Acad Sci USA*, **101**:2174–2178.

Braak, H. and Braak, E., 1991. Neuropathological stageing of alzheimer-related changes. *Acta Neuropathologica*, **82**:239 – 259.

Bulinski, J. C., McGraw, T. E., Gruber, D., Nguyen, H. L., and Sheetz, M. P., 1997. Overexpression of map4 inhibits organelle motility and trafficking in vivo. *Journal of Cell Science*, **110(24)**:3055–3064.

Bunney, W. E., Bunney, B. G., Vawter, M. P., Tomita, H., Li, J., Evans, S. J., Choudary, P. V., Myers, R. M., Jones, E. G., Watson, S. J., *et al.*, 2003. Microarray technology: A review of new strategies to discover candidate vulnerability genes in psychiatric disorders. *Am J Psychiatry*, **160(4)**:657–666.

Canales, R. D., Luo, Y., Willey, J. C., Austermiller, B., Barbacioru, C. C., Boysen, C., Hunkapiller, K., Jensen, R. V., Knight, C. R., Lee, K. Y., *et al.*, 2006. Evaluation of dna microarray results with quantitative gene expression platforms. *Nature Biotechnology*, **24(9)**:1115 – 1122.

Casci, T., 2006. Systems biology: Network fundamentals, via hub genes. *Nature Reviews Genetics*, **7**:664–665.

Catrina, S. B., Okamoto, K., Pereira, T., Brismar, K., and Poellinger, L., 2004. Hyperglycemia regulates hypoxia-inducible factor-1alpha protein stability and function. *Diabetes*, **53(12)**:3226–3232.

Choquette, A. C., Bouchard, L., Houde, A., Bouchard, C., Prusse, L., and Vohl, M. C., 2007. Associations between usf1 gene variants and cardiovascular risk factors in the quebec family study. *Clin Genet*, **71(3)**:245–253.

Christensen, M., Zhou, W., Qing, H., Lehman, A., Philipsen, S., and Song, W., 2004. Transcriptional regulation of bace1, the -amyloid precursor protein beta-secretase, by sp1. *Mol Cell Biol*, **24(2)**:865–874.

Craft, S. and Watson, G. S., 2004. Insulin and neurodegenerative disease: shared and specific mechanisms. *Lancet Neurol*, **3(3)**:169–178.

De Fusco, M., Marconi, R., Silvestri, L., Atorino, L., Rampoldi, L., Morgante, L., Ballabio, A., Aridon, P., and Casari, G., 2003. Haploinsufficiency of atp1a2 encoding the na +/k + pump alpha 2 subunit associated with familial hemiplegic migraine type 2. *Nat Genet*, **33**:192196.

Dennis, Jr, G., Sherman, B. T., Hosack, D. A., Yang, J., Gao, W., Lane, H. C., and Lempicki, R. A., 2003. David: Database for annotation, visualization, and integrated discovery. *Genome Biology*, **4**(5):P3.

Dunckley, T., Beach, T. G., Ramsey, K. E., Grover, A., Mastroeni, D., Walker, D. G., LaFleur, B. J., Coon, K. D., Brown, K. M., Caselli, R., *et al.*, 2006. Gene expression correlates of neurofibrillary tangles in alzheimer's disease. *Neurobiology of Aging*, **27(10)**:1359–1371.

Erbel, P. J., Card, P. B., Karakuzu, O., Bruick, R. K., and Gardner, K. H., 2003. Structural basis for pas domain heterodimerization in the basic helixloophelix-pas transcription factor hypoxia-inducible factor. *Proc Natl Acad Sci U S A*, **100(26)**:15504–15509.

Fang, J., Nakamura, T., Cho, D. H., Gu, Z., and Lipton, S. A., 2007. S-nitrosylation of peroxiredoxin 2 promotes oxidative stress-induced neuronal cell death in parkinson's disease. *Proc Natl Acad Sci U S A*, **104(47)**:18742–18747.

Firulli, B. A., Howard, M. J., McDaid, J. R., McIlreavey, L., Dionne, K. M., Centonze, V. E., Cserjesi, P., Virshup, D. M., and Firulli, A. B., 2003. Pka, pkc, and the protein phosphatase 2a influence hand factor function: a mechanism for tissue-specific transcriptional regulation. *Mol Cell*, **12(5)**:1225–1237.

Germack, R., Leon-Velarde, F., Valdes De La Barra, R., Farias, J., Soto, G., and Richalet, J. P., 2002. Effect of intermittent hypoxia on cardiovascular function, adrenoceptors and muscarinic receptors in wistar rats. *Exp Physiol*, **87(4)**:453–460.

Haapasalo, A., Sipola, I., Larsson, K., Akerman, K., Stoilov, P., Stamm, S., Wong, G., and Castren, E., 2002. Regulation of trkb surface expression by brain-derived neurotrophic factor and truncated trkb isoforms. *J. Biol. Chem*, **277(45)**:43160–43167.

Huber, K., Christ, G., Wojta, J., and Gulba, D., 2001. Plasminogen activator inhibitor type-1 in cardiovascular disease. *Thromb Res*, **103(suppl 1)**:S7–S19.

Irizarry, R. A., Wu, Z., and Jaffee, H. A., 2006. Comparison of affymetrix genechip expression measures. *Bioinformatics*, **22(7)**:789–794.

Janson, J., Laedtke, T., Parisi, J. E., O'Brien, P., Petersen, R. C., and Butler, P. C., 2004. Increased risk of type 2 diabetes in alzheimer disease. *Diabetes*, **53(2)**:474–481.

Karege, F., Perret, G., Bondolfi, G., Schwald, M., Bertschy, G., and Aubry, J. M., 2002. Decreased serum brain-derived neurotrophic factor levels in major depressed patients. *Psychiatry Res*, **109**:143148.

Kimura, R., Kamino, K., Yamamoto, M., Nuripa, A., Kida, T., Kazui, H., Hashimoto, R., Tanaka, T., Kudo, T., Yamagata, H., *et al.*, 2007. The dyrk1a gene, encoded in chromosome 21 down syndrome critical region, bridges between beta-amyloid production and tau phosphorylation in alzheimer disease. *Hum Mol Genet*, **16(1)**:15–23.

Kojima, N. and Shirao, T., 2007. Synaptic dysfunction and disruption of postsynaptic drebrinactin complex: A study of neurological disorders accompanied by cognitive deficits. *Neuroscience Research*, **58(1)**:1–5.

Komulainen, K., Alanne, M., Auro, K., Kilpikari, R., Pajukanta, P., Saarela, J., Ellonen, P., Salminen, K., Kulathinal, S., Kuulasmaa, K., *et al.*, 2006. Risk alleles of usf1 gene predict cardiovascular disease of women in two prospective studies. *PLoS Genet*, **2(5)**:e69.

Krabbe, K., Nielsen, A., Krogh-Madsen, R., Plomgaard, P., Rasmussen, P., Erikstrup, C., Fischer, C., Lindegaard, B., Petersen, A., Taudorf, S., *et al.*, 2007. Brain-derived neurotrophic factor (bdnf) and type 2 diabetes. *Diabetologia*, **50(2)**:431–438(8).

Krapfenbauer, K., Engidawork, E., Cairns, N., Fountoulakis, M., and Lubec, G., 2003. Aberrant expression of peroxiredoxin subtypes in neurodegenerative disorders. *Brain Research*, **967(1)**:152–160.

Laske, C., Stransky, E., Leyhe, T., Eschweiler, G. W., Wittorf, A., Richartz, E., Bartels, M., Buchkremer, G., and Schott, K., 2006. Stage-dependent bdnf serum concentrations in alzheimers disease. *J Neural Transm*, **113**:12171224.

Li, R., Strohmeyer, R., Liang, Z., Lue, L. F., and Rogers, J., 2004. Ccaat/enhancer binding protein delta (c/ebpdelta) expression and elevation in alzheimer's disease. *Neurobiol Aging*, **25(8)**:991–999.

Liang, W. S., Reiman, E. M., Valla, J., Dunckley, T., Beach, T. G., Grover, A., Niedzielko, T. L., Schneider, L. E., Mastroeni, D., Caselli, R., *et al.*, 2008. Alzheimers disease is associated with reduced expression of energy metabolism genes in posterior cingulate neurons. *Proc Natl Acad Sci U S A*, **105(11)**:4441–4446.

MacKnight, C., Rockwood, K., Awalt, E., and McDowell, I., 2002. Diabetes mellitus and the risk of dementia, alzheimer's disease and vascular cognitive impairment in the canadian study of health and aging. *Dement Geriatr Cogn Disord*, **14(2)**:77–83.

Maltepe, E., Schmidt, J. V., Baunoch, D., Bradfield, C. A., and Simon, M. C., 1997. Abnormal angiogenesis and responses to glucose and oxygen deprivation in mice lacking the protein arnt. *Nature*, **386(6623)**:403–407.

Miller, J. A., Oldham, M. C., and Geschwind, D. H., 2008. A systems level analysis of transcriptional changes in alzheimer's disease and normal aging. *J Neurosci*, **28(6)**:1410–1420.

Newman, M. and Girvan, M., 2004. Finding and evaluating community structure in networks. *Phys. Rev. E*, **69(2)**:026113.

Ng, M. C., Miyake, K., So, W. Y., Poon, E. W., Lam, V. K., Li, J. K., Cox, N. J., Bell, G. I., and Chan, J. C., 2005. The linkage and association of the gene encoding upstream stimulatory factor 1 with type 2 diabetes and metabolic syndrome in the chinese population. *Diabetologia*, **48(10)**:2018–2024.

Nguyen, H. L., Gruber, D., McGraw, T., Sheetz, M. P., and Bulinski, J. C., 1998. Stabilization and functional modulation of microtubules by microtubule-associated protein 4. *Biol. Bull.*, **194(3)**:354–357.

Norris, C. M., Kadish, I., Blalock, E. M., Chen, K. C., Thibault, V., Porter, N. M., Landfield, P. W., and Kraner, S. D., 2005. Calcineurin triggers reactive/inflammatory processes in astrocytes and is upregulated in aging and alzheimers models. *J Neurosci*, **25(18)**:46494658.

Pan, Y. S., Lee, Y. S., Lee, Y. L., Lee, W. C., and Hsieh, S. Y., 2006. Differentially profiling the low-expression transcriptomes of human hepatoma using a novel ssh/microarray approach. *BMC Genomics*, **7**:131.

Pastorcic, M. and Das, H. K., 2003. Ets transcription factors er81 and elk1 regulate the transcription of the human presenilin 1 gene promoter. *Brain Res Mol Brain Res*, **113(1-2)**:57–66.

Peers, C., Pearson, H. A., and Boyle, J. P., 2007. Hypoxia and alzheimer's disease. *Essays Biochem*, **43**:153–164.

Pereira, A. C., Wu, W., and Small, S. A., 2007. Imaging-guided microarray: isolating molecular profiles that dissociate alzheimer's disease from normal aging. *Ann N Y Acad Sci*, **1097**:225–238.

Perez-Capote, K., Saura, J., Serratosa, J., and Sola, C., 2006. Expression of c/ebpalpha and c/ebpbeta in glial cells in vitro after inducing glial activation by different stimuli. *Neuroscience Letters*, **410(1)**:25–30.

Qu, D., Rashidian, J., Mount, M. P., Aleyasin, H., Parsanejad, M., Lira, A., Haque, E., Zhang, Y., Callaghan, S., Daigle, M., *et al.*, 2007. Role of cdk5-mediated phosphorylation of prx2 in mptp toxicity and parkinson's disease. *Neuron*, **55(1)**:37–52.

Quandt, K., Frech, K., Karas, H., Wingender, E., and Werner, T., 1995. Matind and matinspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data. *Nucleic Acids Res*, **23(23)**:4878–4884.

Reid, S. J., van Roon-Mom, W. M., Wood, P. C., Rees, M. I., Owen, M. J., Faull, R. L., Dragunow, M., and Snell, R. G., 2004. Tbp, a polyglutamine tract containing protein, accumulates in alzheimer's disease. *Brain Res Mol Brain Res*, **125(1-2)**:120–128.

Ricciarelli, R., d'Abramo, C., Massone, S., Marinari, U., Pronzato, M., and Tabaton, M., 2004. Microarray analysis in alzheimer's disease and normal aging. *IUBMB Life*, **56(6)**:349–354.

Rosendorff, C., Beeri, M. S., and Silverman, J. M., 2007. Cardiovascular risk factors for alzheimer's disease. *Am J Geriatr Cardiol*, **16(3)**:143–149.

Rossler, M., Zarski, R., Bohl, J., and Ohm, T. G., 2002. Stage-dependent and sector-specific neuronal loss in hippocampus during alzheimers disease. *Acta Neuropathol*, **103**:363–369.

Ruan, J. and Zhang, W., 2006a. Identification and evalua- tion of weak community structures in networks. *Proc. National Conf. on AI, (AAAI-06)*, :470–475.

Ruan, J. and Zhang, W., 2008. Identifying network communities with a high resolution. *Physical Review E*, **016104**.

Ruan, J. and Zhang, W., December 2006b. Identification and evaluation of functional modules in gene co-expression networks. *Proc. of RECOMB Satellite Conferences on Systems Biology and Computational Proteomics, San Diego, CA*, .

Salero, E., Gimenez, C., and Zafra, F., 2003. Identification of a non-canonical e-box motif as a regulatory element in the proximal promoter region of the apolipoprotein e gene. *J. Biochem*, **370**:979986.

Sandelin, A., Alkema, W., Engstrom, P., Wasserman, W. W., and Lenhard, B., 2004. Jaspar: an open-access database for eukaryotic transcription factor binding profiles. *Nucleic Acids Res*, **32(Database issue**:D91–94.

Santpere, G., Nieto, M., Puig, B., and Ferrer, I., 2006. Abnormal sp1 transcription factor expression in alzheimer disease and tauopathies. *Neurosci Lett*, **397(1-2)**:30–34.

Shi, J., Xiang, Y., and Simpkins, J. W., 1997. Hypoglycemia enhances the expression of mrna encoding beta-amyloid precursor protein in rat primary cortical astroglial cells. *Brain Research*, **772(1-2)**:247–251.

Shi, J., Zhang, S., Tang, M., Liu, X., Li, T., Han, H., Wang, Y., Guo, Y., Zhao, J., Li, H., *et al.*, 2004. Possible association between cys311ser polymorphism of paraoxonase 2 gene and late-onset alzheimer's disease in chinese. *Brain Res Mol Brain Res.*, **120(2)**:201–204.

Sommerfeld, M. T., Schweigreiter, R., Barde, Y. A., and Hoppe, E., 2000. Down-regulation of the neurotrophin receptor trkb following ligand binding. evidence for an involvement of the proteasome and differential regulation of trka and trkb. *J. Biol. Chem*, **275(12)**:8982–8990.

STAMPFER, M. J., 2006. Cardiovascular disease and alzheimer's disease: common links. *Journal of internal medicine*, **260(3)**:211–223(13).

STEWART, R., 1998. Cardiovascular factors in alzheimer's disease. *J Neurol Neurosurg Psychiatry*, **65**:143–147.

Sun, X., He, G., Qing, H., Zhou, W., Dobie, F., Cai, F., Staufenbiel, M., Huang, L. E., and Song, W., 2006. Hypoxia facilitates alzheimer's disease pathogenesis by up-regulating bace1 gene expression. *Proc Natl Acad Sci U S A*, **103(49)**:18727–18732.

Tomita, S., Fujita, T., Kirino, Y., and Suzuki, T., 2000. Pdz domain-dependent suppression of nf-kappa b/p65-induced abeta 42 production by a neuron-specific x11-like protein. *J Biol Chem*, **275(17)**:13056–13060.

Tong, L., Balazs, R., Thornton, P. L., and Cotman, C. W., 2004. Beta-amyloid peptide at sublethal concentrations downregulates brain-derived neurotrophic factor functions in cultured cortical neurons. *J Neurosci*, **24(30)**:6799–809.

Tsai, S. J., 2003. Brain-derived neurotrophic factor: a bridge between major depression and alzheimer's disease? *Med Hypotheses*, **61(1)**:110–113.

Tsuda, H., Jafar-Nejad, H., Patel, A. J., Sun, Y., Chen, H. K., Rose, M. F., Venken, K. J., Botas, J., Orr, H. T., Bellen, H. J., *et al.*, 2005. The axh domain of ataxin-1 [?] mediates neurodegeneration through its interaction with gfi-1/senseless proteins. *Cell*, **122(4)**:633–644.

Tusher, V. G., Tibshirani, R., and Chu, G., 2001. Significance analysis of microarrays applied to the ionising radiation response. *Proc Natl Acad Sci USA*, **98**:5116 – 5121.

Tyler, W. J., Alonso, M., Bramham, C. R., and Pozzo-Miller, L. D., 2002. From acquisition to consolidation: on the role of brain-derived neurotrophic factor signaling in hippocampal-dependent learning. *Learn Mem*, **9**:224–237.

Wakaguri, H., Yamashita, R., Suzuki, Y., Sugano, S., and Nakai, K., 2008. Dbtss: database of transcription start sites, progress report 2008. *Nucleic Acids Res*, **36(Database issue)**:D97–101.

Wang, G. and Zhang, W., 2006. A steganalysis-based approach to comprehensive identification and characterization of functional regulatory elements. *Genome Biology*, **7(6):R49**.

Winsky-Sommerer, R., Grouselle, D., Rougeot, C., Laurent, V., David, J. P., Delacourte, A., Dournaud, P., Seidah, N. G., Lindberg, I., Trottier, S., *et al.*, 2003. The proprotein convertase pc2 is involved in the maturation of prosomatostatin to somatostatin-14 but not in the somatostatin deficit in alzheimer's disease. *Neuroscience*, **122(2)**:437–447.

Y, M., M, P., B, M., J, L., C, Z., M, D. J., A, O. J., I, F. M., MK, O., J, T. A., *et al.*, 2001. Inflammatory responses to amyloidosis in a transgenic mouse model of alzheimers disease. *American Journal of Pathology*, **158**:1345–1354.

Yamada, K., Mizuno, M., and Nabeshima, T., 2002. Role for brain-derived neurotrophic factor in learning and memory. *Life Sci*, **70**:735–744.

Yelon, D., Ticho, B., Halpern, M. E., Ruvinsky, I., Ho, R. K., Silver, L. M., and Stainier, D. Y., 2000. The bhlh transcription factor hand2 plays parallel roles in zebrafish heart and pectoral fin development. *Development*, **127(12)**:2573–2582.

Yue, H., Eastman, P. S., Wang, B. B., Minor, J., Doctolero, M. H., Nuttall, R. L., Stack, R., Becker, J. W., Montgomery, J. R., Vainer, M., *et al.*, 2001. An evaluation of the performance of cdna microarrays for detecting changes in global mrna expression. *Nucleic Acids Res*, **29(8)**:E41–1.

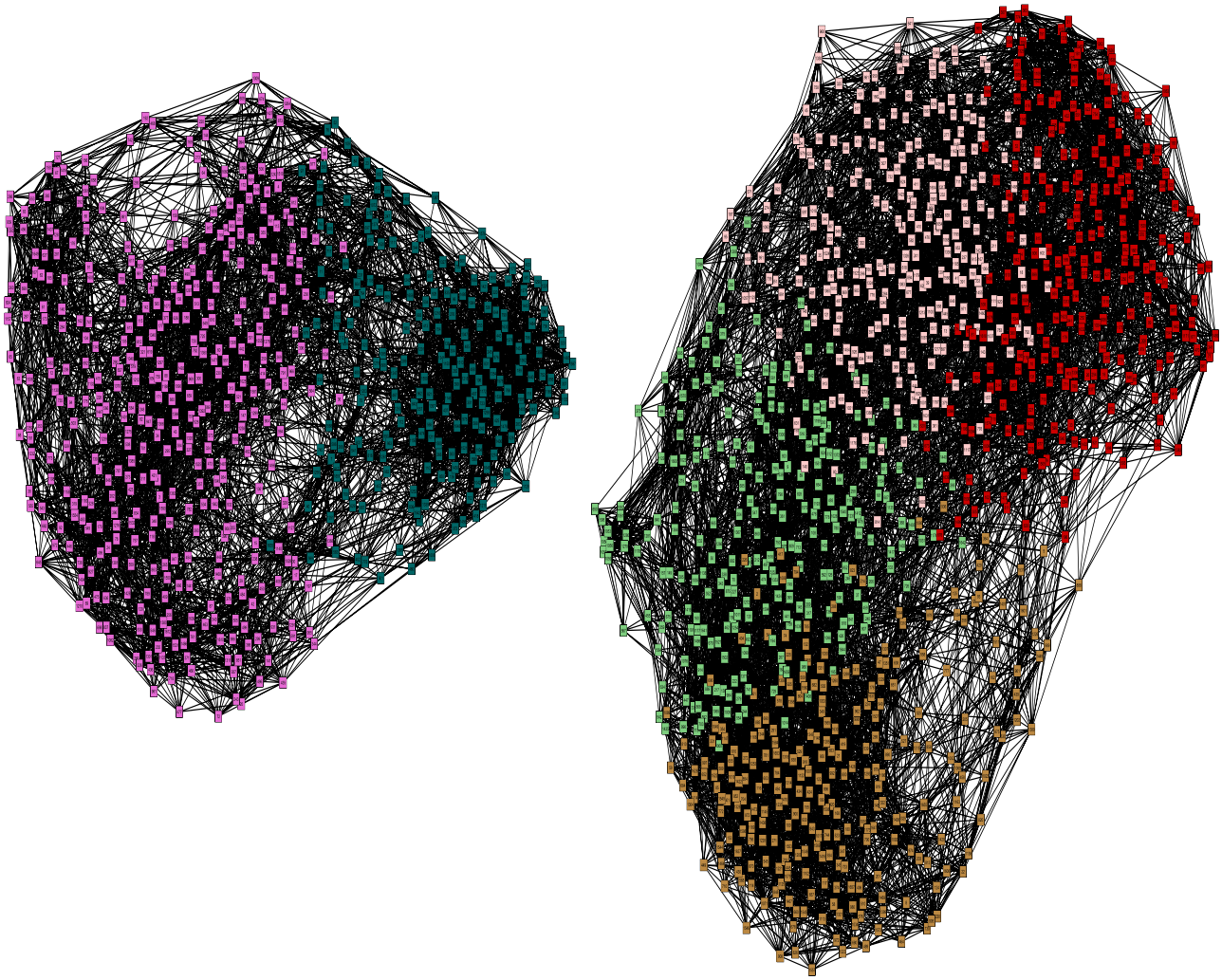

Figure 3: Co-expression network with 6 modules. A node refers to a gene and the weight of an edge is the Pearson correlation coefficient between expression profiles of a pair of genes scaled to within [0,1]. The 2 large groups are two sets of genes with anti-correlated expression patterns. The group on the left contain all upregulated genes and the group on the right consist of downregulated genes. The length of each edge and the position of each node/module does not have any biological meaning and are arbitrarily chosen for proper visualisation. Best viewed in colour.