Journal of Proteomics & Bioinformatics - Open Access

www.omicsonline.com JPB/Vol.2/December 2009

Research Article OPEN ACCESS Freely available online doi:10.4172/jpb.1000113

Computational Identification of Alzheimer's Disease Specific Transcription Factors using Microarray Gene Expression Data

Vishalini Krishnamurthy, Nicy Sweety Issac and Jeyakumar Natarajan*

Data and Text Mining Laboratory, Department of Bioinformatics, Bharathiar University, Coimbatore 641046, India

Abstract

Alzheimer's disease is the most common form of dementia affecting millions of older people world wide. Identification of transcriptional factor binding sites of disease specific co-expressed genes and the possible transcriptional regulation of the genes will lead to a better understanding of complex diseases such as Alzheimer's disease. However, the regulatory mechanisms driving these changes, in particular the networks of transcription factors involved, is not fully understood to date. The computational identification of conserved TFBS in the regulatory regions of hundreds of genes at a time especially suited for microarray gene expression datasets. We report clusters of co-expressed genes and the identification of conserved TFBSs using microarray gene expression data sets. We investigated microarray gene expression data from Gene Expression Omnibus (GEO) specific to Alzheimer's disease. The dataset consists of 14 normal and 14 Alzheimer disease samples. Differential expression analysis results 240 differentially expressed genes which are more significant. Hierarchical clustering of these significance genes shows eight clusters of co-expressed genes. The detection of over-represented transcription factor binding sites in the promoters regions of co-expressed genes reveals transcription factor binding site classes ZEB1, MZF1 1-4, ZNF354C, ELF5 and SPIB in upstream of human promoter and responsible for apoptosis.

Keywords: Alzheimer's disease; Microarray; Transcription factor; Microarray tools

Introduction

Alzheimer's disease (AD) also called Alzheimer disease is a complex progressive neurodegenerative disorder of the brain and most common form of age-related cognitive impairment (Tiraboschi et al., 2004). The cause and progression of Alzheimer's disease are not yet well understood. However, the ongoing research indicates that the disease is associated with plaques and tangles in the brain (Dunckley et al., 2006). AD is characterized by two pathologic hallmark lesions that consist of extracellular plaques of amyloid-beta peptides and intracellular neurofibrillary tangles composed of hyperphosphorylated microtubular protein tau (Okuizumi and Tsuji, 2007) (Figure 1). Recent advances in molecular genetics have enabled the identification of the causative genes for Alzheimer's disease and the most common forms of AD are considered to be polygenic disorders (Blalock et al., 2005). AD poses a great challenge to patients, oncologists, and biologists due to polygenic disorders and the involvement of large number of genes and their complex interactions.

Since microarray technology allows massively parallel analysis of most genes expressed in a tissue it has become a popular gene expression screening tool in the molecular investigation of polygenic disease such as AD (Blalock et al., 2004; Kong et al., 2009; Pasinetti, 2001). Microarray technology today is rapidly uncovering patterns of genetic activity and showing insight into prediction of gene functions (Pan, 2006; Li et al., 2006; Tenenbaum et al., 2008), pathways (Manoli et al., 2006; Veerla and Höglund, 2006) and transcription factor binding sites (TFBSs) (Park et al., 2002; Haverty et al., 2004). The challenges that lie here include systematically identifying the functions of all AD associated genes, and continuing the efforts to decipher their pathways and regulatory networks. This information will help to understand the mechanism of AD development and assist in the identification of effective therapeutic targets for disease control and eradication.

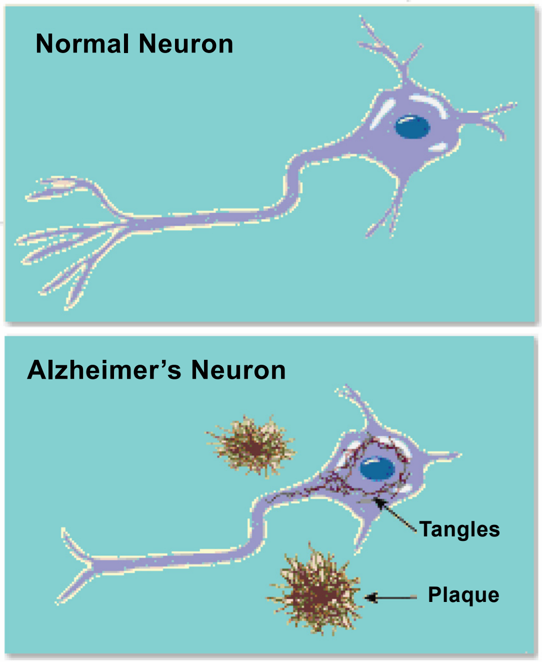

Figure 1: Differentiation of normal and Alzheimer's neuron. **A:** normal neuron without Alzheimer's. **B:** Deposition of neurofibrillary tangles and amyloid plaques in the nerve cells of brain with Alzheimer's disease.

***Corresponding author:** N. Jeyakumar, Data and Text Mining Laboratory, Department of Bioinformatics, Bharathiar University, Coimbatore 641046, India, E-mail: n.jeyakumar@yahoo.co.in

Received November 05, 2009; **Accepted** December 18, 2009; **Published** December 20, 2009

Citation: Krishnamurthy V, Issac NS, Natarajan J (2009) Computational Identification of Alzheimer's Disease Specific Transcription Factors using Microarray Gene Expression Data. J Proteomics Bioinform 2: 505-508. doi:10.4172/jpb.1000113

The transcription factor binding sites (TFBSs) 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 CVDs (cardiovascular diseases), stroke and diabetes (Tavazoie et al., 1999). Gene-expression microarrays have been analyszed using clustering algorithms that group genes and samples on the basis of expression profiles, and statistical methods that score genes on the basis of their relevance to various clinical attributes (Ray, 2008). Transcription factors act as critical molecular switches in promoting neuronal survival (Burton et al., 2002).

In this work, we performed a microarray based study of a dataset consists of 14 normal and 14 Alzheimer's disease samples. We first used microarray expression profiling to distinguish the broadest set of genes that showed differential expression levels across two disease types normal vs. AD. Second, we clustered the differentially expressed genes, based on their expression profiles, into sets of putatively co-regulated genes. Finally, we attempted to identify the transcription factors, as well as their corresponding binding sites, which regulate the observed expression differences of the genes in the differentially and co-expressed gene set. . As gene regulators are important targets to treat diseases, we have identified TFBSs ZEB1, MZF1 1-4, ZNF354C and SPIB that would have a high therapeutically value to treat Alzheimer's disease.

Materials and Methods

The systematic identification and characterization of Alzheimer's disease specific transcription factors using microarray gene expression data is illustrated in Figure 2.

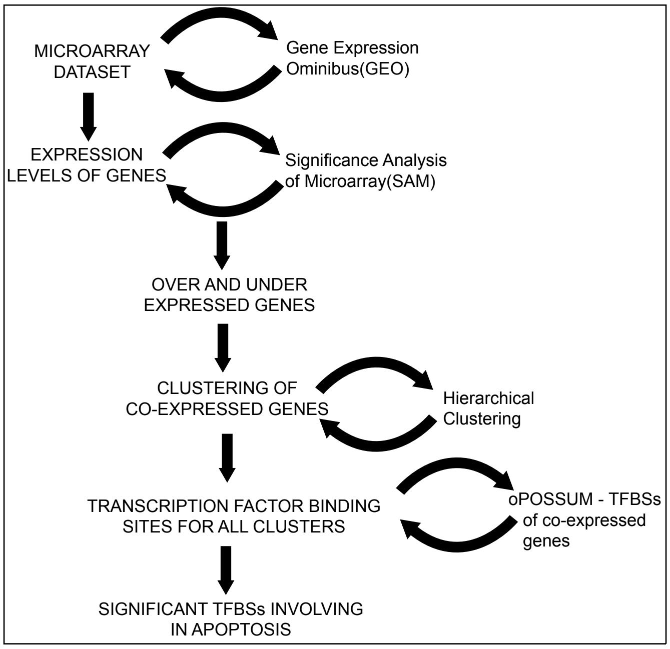

Figure 2: Methodology. Microarray dataset is obtained from Gene Expression Omnibus (GEO) in which the expression levels of each gene are present for 28 different samples. The differentially expressed genes which are most significant are identified through Significance Analysis of Microarray (SAM). Based upon the co-expression, the genes are clustered using hierarchical clustering viewed in tree form. Transcription factor binding sites for all the clusters of co-expressed genes are identified using oPOSSUM. Predicting the significant and common TFBSs.

Alzheimer's gene microarray data

The dataset of Maes et al., (2007) consists of 14 normal controls and 14 AD affected samples obtained from Gene Expression Omnibus (GEO Accession Number: GDS2601) was used in this study. Gene expression was measured using GPL1211: NIA MGC, Mammalian Genome Collection (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL1211) covering 9601 genes for AD (14 samples) and control (14 samples).

Differential gene expression

The gene expression of control and AD stage has been considered and compared for the identification of differentially expressed genes in AD stage. Significance analysis of Microarray (SAM) determines the significant changes in expression of genes between different biological stages based on statistical analysis of modified gene specific t-test (Tusher et al., 2001).

Clustering of co-expressed genes

After selecting the differentially expressed genes, we clustered the genes based on the expression level of genes to find the co-expressed gene clusters. We used MultiExperiment Viewer (MEV) software package from TIGR (Saeed et al., 2003) for hierarchical clustering of microarray data, using Euclidean Distance metrics and Average Linkage Clustering algorithms. The Tree View (supplementary materiel 1) shows the relationship between the genes based on the gene expression profile.

Analysis of enrichment for TFBS

Each clusters of gene set was analyzed for enrichment of TFBS using the oPOSSUM program (Ho Sui et al., 2007). The conserved non-coding regions of the promoters were searched for matches to all TFBS profiles in the JASPAR (Sandelin et al., 2004) database. For each transcript, the top 10% of conserved regions in the 2000-bp upstream/downstream sequences between mouse and humans with minimum conservation of 70% and matrix match threshold of 80% was scanned for TFBS using a position weight matrices algorithm.

Results and Discussion

The results obtained through Significance Analysis of Microarrays (SAM) (Saeed et al., 2003) reveals that out of 25,577 genes in the microarray data set 240 genes were identified as differentially expressed genes between AD samples and controls at a false discovery rate of 0.1%. We then identified the groups of co-regulated genes based on the hypothesis that genes with strongly correlated mRNA expression profiles are more likely to have their promoter regions bound by a common transcription factor (Allocco et al., 2004). We used hierarchal clustering of differentially expressed genes and identified eight groups of co-expressed genes clusters. A figure showing the eight clusters of co-expressed genes is provided in Additional data file 1.

Next, we identified the conserved TFBSs among co-expressed genes clusters. We used Entrez-gene to define the transcription start sites (TSS) and determined overrepresented transcription factor binding sites in the promoter regions. Each clusters of gene set was analyzed for enrichment of TFBS using the oPOSSUM program (Ho Sui et al., 2007). Transcription factors ZEB1, MZF1 1-4, ZNF354C, ELF5 and SPIB were found to have their binding sites in most of the genes which are differen-

Cluster Number	Number of genes in clusters	Significant TFBS present in the each cluster
1	24	MZF1 1-4, ZNF354C, ZEB1, ELK5,SPIB
2	49	ZEB1, MZF1 1-4, ELF5, ZNF354, SPIB
3	38	MZF1 1-4, ZNF354C, ZEB1, ELF5, SPIB
4	26	SPIB, ZEB1, ZNF354C, MZF1 1-4, SOX5, ELF5
5	24	MZF1 1-4, ELF5, ZEB1, ZNF354C, SPIB
6	18	SPIB, ZEB1, MZF1 1-4, ZNF354C, ELF5
7	28	ZEB1, MZF1 1-4, ZNF354C, SPIB, ELF5
8	28	ELF5, ZNF354C, ZEB1, SPIB, MZF1 1-4

Table 1: Total number of clusters, number of genes in each cluster and signifigant transcription factors.

tially expressed in AD. The enriched TFBS for co-expressed expressed genes in AD was illustrated in Table 1 and the corresponding binding sites from JASPAR (Sandelin et al., 2003) database was illustrated in Figure 3.

On further literature analysis of these common transcription factor binding sites we found that ZEB1, Zinc finger/homeodomain serve as DNA binding domain with greater affinity for a subset of E box and E-box- like sequences (CACCTG). ZEB1/zfh-1 transcriptional repressor regulates muscle differentiation and expressed in central nervous system (Antonio and Douglas, 2000). A study by Schmalhofer et al., (2009) reveals the molecular interconnection of ZEB1 with E-cadherin, β-catenin, and WNT signaling in cancerogenesis. WNT signaling regulates dendrite morphogenesis. Dendritic pathology and decrease of dendritic spine density are prominent phenomena in early cases of AD. ZEB1, through WNT signaling would have a role in the dendritic degeneration in AD (Baloyannis, 2009). GATA-3 is essential for T-cell development and a recent study by Dontje et al. suggests that Spi-B TF is a key regulator of Dentritic cells development (Dontje et al., 2006). T-cell development is inhibited by Spi-B through induction of apoptosis in T-cell precursors without inhibiting the differentiation (Schotte et al., 2003). T cell population is a response to the presence of Ameloid β aggregates in AD.

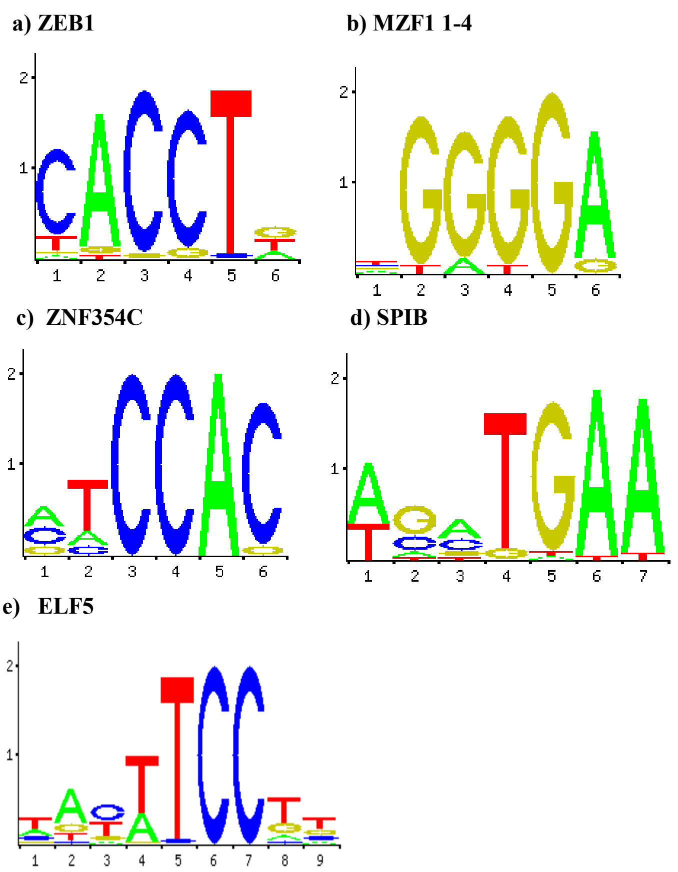

Figure 3: Representation of transcription factor binding sites from JASPAR database for the common transcription factor (a) ZEB1, (b) MZF1 1-4, (c) ZNF354C, (d) SPIB,(e) ELF5.

Leucine zipper down-regulated in cancer (LDOC1), is a gene that encodes a leucine-zipper protein characteristic for early-phase apoptotic events and reduced cell viability in human cell lines. Another transcription factor, MZF1, interacts with LDOC1 and enhances the activity of LDOC1 for inducing apoptosis (Inoue et al., 2005). MZF1 was found to play a key role in cell lines representing early stages of myeloid differentiation and derivation of ES cell lines involved in growth, differentiation, and apoptosis (Dong et al., 2008).

ETS transcription factors ELF5 is essential for developmental processes in the embryo and in the mammary gland during pregnancy (Oakes et al., 2006). ETS factors involving in early embryonic development ELF5 modulate the expression of a variety of genes involved in various cellular processes, including cell proliferation, differentiation and apoptosis (Jedlicka and Gutierrez-Hartmann, 2008).

The literature information clearly indicates that these common transcription factor binding sites (ZEB1, MZF1 1-4, ZNF354C, ELF5 and SPIB) are the regulator of Alzheimer's disease during apoptosis pathway and inducing cell death and apoptosis.

Conclusion

One of the challenges of computational biology is to identify genomic binding sites for transcription factors and the direct downstream targets they affect. Identification of such binding sites would allow the development of more accurate gene networks, interactions and an understanding of important biological pathways. As a starting point for that we described an analysis using public microarray experiments in AD. We identified groups of genes, or co-expressed modules, that undergo similar changes in expression and identified conserved TFBSs which are considered as the gene regulators for Alzheimer's disease. As gene regulators are important targets to treat diseases, the identified TFBSs ZEB1, MZF1 1-4, ZNF354C, ELF5 and SPIB would have a high therapeutically value to treat Alzheimer's disease.

References

1. Allocco DJ, Kohane IS, Butte AJ (2004) Quantifying the relationship between co-expression, co-regulation and gene function. BMC Bioinformatics 5: 18. » CrossRef » PubMed » Google Scholar
2. Antonio AP, Douglas CD (2000) Differential expression and function of members of the zfh-1 of zinc finger/homeodomain repressors. PNAS 97: 6391-6396. » CrossRef » PubMed » Google Scholar
3. Baloyannis S (2009) Dendritic pathology in Alzheimer's disease. Neurological Sciences 283: 153-157. » CrossRef » PubMed » Google Scholar
4. Blalock EM, Chen KC, Stromberg AJ, Norris CM, Kadish I, et al. (2005) Harnessing the power of gene microarrays for the study of brain aging and Alzheimer's disease: statistical reliability and functional correlation. Ageing Res Rev 4: 481-512. » CrossRef » PubMed » Google Scholar
5. 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. PNAS 101: 2173-2178. » CrossRef » PubMed » Google Scholar
6. Burton TR, Dibrov A, Kashour T, Amara FM (2002) Anti-apoptotic wild-type Alzheimer amyloid precursor protein signaling involves the p38 mitogen-activated protein kinase/MEF2 pathway. Brain Res Mol Brain Res 108: 102-120. » CrossRef » PubMed » Google Scholar
7. Dong S, Ying S, Kojim T, Shiraiw M, Kawada A, et al. (2008) Crucial Roles of MZF1 and Sp1 in the Transcriptional Regulation of the Peptidylarginine Deiminase Type I Gene (PADI1) in Human Keratinocytes. J Invest Dermatol 128: 549-557. » CrossRef » PubMed » Google Scholar
8. Dontje W, Schotte R, Cupedo T, Nagasawa M, Scheeren F, et al. (2006) Delta-like1–induced Notch1 signaling regulates the human plasmacytoid dendritic cell versus T-cell lineage decision through control of GATA-3 and Spi-B. Immunobiology 107: 2446-2452. » CrossRef » PubMed » Google Scholar
9. 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 10: 1359-1371. » CrossRef » PubMed » Google Scholar
10. Haverty PM, Hansen U, Weng Z (2004) Computational inference of transcriptional regulatory networks from expression profiling and transcription factor binding site identification. Nucleic Acids Res 32: 179-88. » CrossRef » PubMed » Google Scholar
11. Ho Sui SJ, Fulton LD, Arenillas DJ, Kwon AT, Wasserman WW (2007) oPOSSUM: integrated tools for analysis of regulatory motif over-representation. Nucleic Acids Res 35: W245-52. » CrossRef » PubMed » Google Scholar
12. http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL1211.
13. Inoue M, Takahashi K, Niide O, Shibata M, Fukuzawa M (2005) LDOC1, a novel MZF-1-interacting protein, induces apoptosis. FEBS Lett 579: 604-608. » CrossRef » PubMed » Google Scholar
14. Jedlicka P, Gutierrez-Hartmann A (2008) Ets transcription factors in intesti nal morphogenesis, homeostasis and disease. Histol Histopathol 23: 1417-1424. » CrossRef » PubMed » Google Scholar
15. Kong W, Mou X , Liu Q, Chen Z, Vanderburg CR, et al. (2009) Independent component analysis of Alzheimer's DNA Microarray gene expression data. Mol Neurodegener 4: 5. » CrossRef » PubMed » Google Scholar
16. Li XL, Tan YC, Ng SK (2006) Systematic gene function prediction from gene expression data by using a fuzzy nearest-cluster method. BMC Bioinformatics 7: S23. » CrossRef » PubMed » Google Scholar
17. Manoli T, Gretz N, Grone HJ, Kenzelmann M, Eils R, et al. (2006) Group testing for pathway analysis improves comparability of different microarray data sets. Bioinformatics Advance Access. » CrossRef » PubMed » Google Scholar
18. Maes OC, Xu S, Yu B, Chertkow HM, Wang E, et al. (2007) Transcriptional profiling of Alzheimer blood mononuclear cells by microarray. Neurobiol Aging 28: 1795-809. » CrossRef » PubMed » Google Scholar
19. Oakes SR, Hilton HN, Garvan CJO (2006) Key stages in mammary gland development - The alveolar switch: coordinating the proliferative cues and cell fate decisions that drive the formation of lobuloalveoli from ductal epithelium. Breast Cancer Res 8: 207. » CrossRef » PubMed » Google Scholar
20. Okuizumi K, Tsuji S (2007) Alzheimer's disease as a polygenic disease. Neuropathology 18: 111-115. » CrossRef » PubMed » Google Scholar
21. Pan W (2006) Incorporating gene functions as priors in model-based clustering of microarray gene expression data. Bioinformatics 22: 795-801. » CrossRef » PubMed » Google Scholar
22. Park PJ, Butte AJ, Kohane IS (2002) Comparing expression profiles of genes with similar promoter regions. Bioinformatics 18: 1576-84. » CrossRef » PubMed » Google Scholar
23. Pasinetti GM (2001) Use of cDNA microarray in the search for molecular markers involved in the onset of Alzheimer's disease dementia. J Neurosci Res 6: 471-476. » CrossRef » PubMed » Google Scholar
24. 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. » CrossRef » PubMed » Google Scholar
25. Saeed AI, Sharov V, White J, Li J, Liang W, et al. (2003) TM4: a free, open-source system for microarray data management and analysis. Biotechniques 34: 374-8. » CrossRef » PubMed » Google Scholar
26. Sandelin A, Alkema W, Engstrom P, Wasserman WW, Lenhard B (2004) JASPAR: an open-access database for eukaryotic transcription factor binding profiles. Nucleic Acids Res D91-94. » CrossRef » PubMed » Google Scholar
27. Schmalhofer O, Brabletz S, Brabletz T (2009) E-cadherin, β-catenin, and ZEB1 in malignant progression of cancer. Cancer Metastasis Rev 28: 151-166. » CrossRef » PubMed » Google Scholar
28. Schotte R, Rissoan MC, Bendriss-Vermare N, Bridon JM, Duhen T, et al. (2003) The transcription factor Spi-B is expressed in plasmacytoid DC precursors and inhibits T-, B-, and NK-cell development. Blood 101: 1015-1023. » CrossRef » PubMed » Google Scholar
29. Tavazoie S, Hughes JD, Campbell MJ, Cho RJ, Church GM (1999) Systematic determination of genetic network architecture. Nature Genet 22: 281-285. » CrossRef » PubMed » Google Scholar
30. Tenenbaum JD, Walker MG, Utz PJ, Butte AJ (2008) Expression-based Pathway Signature Analysis (EPSA): Mining publicly available microarray data for insight into human disease. BMC Med Genomics 1: 51. » CrossRef » PubMed » Google Scholar
31. Tiraboschi P, Hansen LA, Thal LJ, Corey-Bloom J (2004) The importance of neuritic plaques and tangles to the development and evolution of AD. Neurology 62: 1984-9. » CrossRef » PubMed » Google Scholar
32. Tusher VG, Tibshirani R, Chu G (2001) Significance analysis of microarrays applied to the ionizing radiation response. PNAS 98: 5116-5121. » CrossRef » PubMed » Google Scholar
33. Veerla S, Höglund M (2006) Analysis of promoter regions of co-expressed genes identified by microarray analysis. BMC Bioinformatics 7: 384. » CrossRef » PubMed » Google Scholar