Synergies of genes in Alzheimer's disease

Isabel A. Nepomuceno-Chamorro, Jesús S. Aguilar-Ruiz

ETSII, Dpto. Lenguajes y Sistemas Informáticos, Universidad de Sevilla, Spain
School of Engineering, Pablo de Olavide University, Seville, Spain
{inepomuceno}@us.es
{aguilar}@upo.es

1 Introduction

Nowadays, Alzheimer's disease (AD) is the most common form of dementia affecting more than 25 million individuals worldwide. AD is a complex progressive neurodegenerative disorder of the brain and the prevalence of Alzheimer's disease is expected to rise over the next decades [1]. Due to its poligenic nature, unconventional strategies for elucidating the genetic mechanisms and genetic risk factors are necessary.

The microarray technology has made possible the monitoring of hundreds of genes simultaneous. Most studies on analysis of microarray data are based on the identification of differentially expressed genes on which further analysis is performed. There are several approaches to identify differentially expressed genes as feature selection method. However, due to the AD's polygenic nature, it is believed that AD is caused not by defect in a few genes, but rather by variations in a large number of genes and their complex interplay. Therefore, a systems-level approach can provide new insights into the interplay of genes and it is an effective alternative for analyzing this complex disease.

2 Method

A systems biology approach based on the inference of gene association networks from gene expression profiles is used. The methodology, named RegNet , is based on model trees as a method to identify gene interaction networks and it favours localized similarities over more global similarity, which is one of the major drawbacks of correlation-based methods. The methodology consists of three steps. The first step involves building M5' trees. M5' is a model tree algorithm, an extension of regression tree algorithms, which builds several linear models, each one of them built in a leaf of the tree. The aim of the second step is to obtain a set of genes associated to other genes from their prediction ability by means of linear regression functions. Finally, the third step involves learning a graph model of gene-gene associations by assessing the significance of the set of hypothetical evidences. This method is reported in [2] and it is part of a network-based prognostic model reported in [3].

3 Results and discussion

In this experiment we use the data set of [4]. It consists of 14 control and 19 Alzheimer's disease (AD) brain tissue samples. The 14 control samples are Braak stages 0-II with average age 80.1 years and 19 AD affected are Braak stages III-IV with average age 84.7 years (incipient AD). This dataset includes 35722 gene probesets. We run our method on a subset of them which corresponds to 1663 genes obtained from that dataset using a preprocessing step described in [5].

Using David tool, we observed the association of genes with various disease: cardiovascular disease, alzheimer disease and type 2 diabetes. In an enrichment analysis using GO we detected several genes involved in immune, inflammatory or defense response. In the study of Miller et al. [6], 558 transcripts which are common to AD and ageing were identified. We can assert that we found more overlapping set of genes between our results and Miller's study than expected by chance. In the study of Kong et al. [7], they identified 87 significant genes in severe AD representing immunity-related protein, metal-related protein, lipoprotein, neuropeptide and ribosomal protein. We study the overlapping between those significant genes in severe AD and our results obtained with incipient AD. In the input data set, an overlapping of 29 genes can be found between the list of 1663 and those 87 genes. Of these 29 genes, 12 genes were present in the major subgraph (456 genes) of our analysis (greater than expected by chance $p - value = 0.004$). It is worth noting that in these 12 genes we found genes Colrf115 and C20orf149 which are related with ribosomal protein, this genes are significant because impairments in protein synthesis occur in the earliest stages of AD (see Ding et al., 2005). We also found IFITM2 inflammation-related gene, neuroinflammation is believed to be a culprit in AD pathogenesis (see Kong et al., 2009). Finally, we found CHGB, MTF1 and MT1M genes related to metal protein. In the literature, we can observe that the level of metal ion metabolism is closely associated with AD (see Soderling et al., 2001).

References

1. Orevsivc, M., Lotjonen, J., Soininen, H.: Systems medicine and the inte- gration of bioinformatic tools for the diagnosis of Alzheimer's disease. Genome Medicine, 11,83 (2010)
2. Nepomuceno-Chamorro, I., Aguilar-Ruiz, J., Riquelme-Santos, J.: Inferring gene regression networks with model trees. BMC Bioinformatics, 11, 517 (2010)
3. Nepomuceno, I., Azuaje, F. et al.: Prognostic transcriptional association networks: a new supervised approach based on regression trees. Bioinformatics, 27, 252–258 (2011)
4. Dunckley, T., et al.: Gene expression correlates of neurofibrillary tangles in Alzheimers disease. Neurobiol Aging, 27, 1359–1371 (2006)
5. Ray, M., Ruan, J., Zhang, W.: Variations in the transcriptome of alzheimer's disease reveal molecular networks involved in cardiovascular diseases. Genome Biology, 9, R148 (2008)

6. Miller, J., Oldham, M., Geschwind, D.: A Systems Level Analysis of Transcriptional Changes in Alzheimer's Disease and Normal Aging. Journal of Neuroscience, 28, 1410–1420 (2008)
7. Kong,W., Mou,X., Liu,Q., Chen,Z., Vanderburg,C.R., Rogers,J.T., Huang,X.:Independent component analysis of Alzheimer's DNA microarray gene expression data. Molecular Neurodegeneration, 4, 5, (2009)