Rule-based Data Mining Analysis for Microarray Gene Expression Databases in Alzheimer's Disease

Prerna Dua, PhD

Louisiana Tech University Department of Health Informatics and Information Management

Abstract

Alzheimer's disease is a human brain disease that is progressive, leads to profound cognitive impairment, and is always fatal. Approximately 4 million Americans are affected by Alzheimer's disease and the incidence of Alzheimer's disease is expected to increase in the near future. Although Alzheimer's disease research is intensive, the root cause of Alzheimer's disease is not known, and there is widespread general disagreement among many domestic and international research laboratories on theories that may explain disease etiology. One solid approach is to investigate the expression levels of brain-specific genes both in Alzheimer's disease and in normally aging human brain.

The amygdala, hippocampus, and cingular cortex of Alzheimer patients show signs of significant changes in physiological function that accompany the amyloid plaques and neurofibrillary tangles that are hallmarks of the disease. A substantial portion of post-transcriptional gene regulation is controlled by microRNA (miRNA) networks and hence an alteration in miRNA expression is emerging as a significant contributing factor to Alzheimer's disease. Specific miRNA abundance is significantly altered in neurological disorders such as Alzheimer's disease compared with age-matched controls. It is important to discover the biologically significant correlations among co-regulated miRNAs that have a substantial role in the progression of Alzheimer's. MiRNA-146a is found in increased amounts in stressed human brain cells and in Alzheimer's disease, and it plays a crucial role in the regulation of inflammation. Previous work observed that miRNA-146a appears to reduce the amount and bioavailability of Compliment Factor H (CFH), promoting brain cell inflammation and contributing to the development of Alzheimer's disease. In this study, we investigate DNA microarray gene expression data from AD hippocampal tissue of diseased and age-matched controls, specifically concentrating on discovery of discriminatory NF-kappa-B-sensitive patterns with other co-regulated genes using an information theory approach. We developed an unsupervised clustering algorithm based on the K-local hyperplane distance nearest-neighbor algorithm. We further identify the frequently occurring sets of genes by exploiting the associative dependencies and validate our experiments by conducting a biomedical literature search in addition to the known classification schemes to evaluate the efficacy of the selected genes.