Journal of Theoretical Biology 334 (2013) 109–121

Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

Computational studies on Alzheimer's disease associated pathways and regulatory patterns using microarray gene expression and network data: Revealed association with aging and other diseases

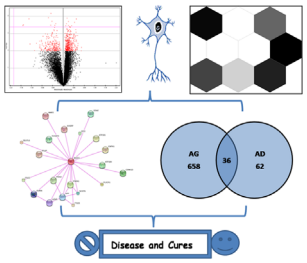
Priya P. Panigrahi, Tiratha Raj Singh *

Department of Biotechnology and Bioinformatics, Jaypee University of Information Technology, Waknaghat, Solan, Himachal Pradesh 173234, India

HIGHLIGHTS

- The association of novel genes and variants for their interaction with AD.
- TFBS identified as a mediocre biological process for AD and AG.
- Physico-chemical analysis for TFBS revealed novel associations.
- Novel information for network motifs such as BiFan, MIM, SIM, and others.
- Unique miRNA targets such as LDB2, and DOPEY1 as a regulatory process for AD.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 29 October 2012
Received in revised form
7 June 2013
Accepted 10 June 2013
Available online 26 June 2013

Keywords:
Enrichment analysis
TFBS
Mirna
Network motifs
Nucleosomes

ABSTRACT

Alzheimer's disease (AD), which is one of the most common age-associated neurodegenerative disorders, affects millions of people worldwide. Due to its polygenic nature, AD is believed to be caused not by defects in single genes, but by variations in a large number of genes and their complex interactions, which ultimately contribute to the broad spectrum of disease phenotypes. Extraction of insights and knowledge from microarray and network data will lead to a better understanding of complex diseases. The present study aimed to identify genes with differential topology and their further association with other biological processes that regulate causative factors for AD, ageing (AG) and other diseases. Our analysis revealed a common sharing of important biological processes and putative candidate genes among AD and AG. Some significant novel genes and other variants for various biological processes have been reported as being associated with AD, AG, and other diseases, and these could be implicated in biochemical events leading to AD from AG through pathways, interactions, and associations. Novel information for network motifs such as BiFan, MIM (multiple input module), and SIM (single input module) and their close variants has also been discovered and this implicit information will help to improve research into AD and AG. Ten major classes for TFs (transcription factors) have been identified in our data, where hundreds of TFBS patterns are being found associated with AD, and other disease. Structural and physico-chemical properties analysis for these TFBS classes revealed association of biological processes involved with other severe human disease. Nucleosomes and linkers positional information could provide insights into key cellular processes. Unique miRNA (micro RNA) targets were identified as another regulatory process for AD. The association of novel genes and variants of existing genes have also been explored for their interaction and association with other diseases that are either directly or indirectly implicated through AG and AD.

* Corresponding author. Tel.: +91 1792 239385; fax: +91 1792 245362.
E-mail address: tiratharaj@gmail.com (T.R. Singh).

0022-5193/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.jtbi.2013.06.013

1. Introduction

AD affects millions of people worldwide and is one of the most common age associated neurodegenerative disorders (Chou, 2004). AD is characterized by a progressive decline in memory associated with other cognitive deficits: judgment, abstraction, language, attention and visuoconstructive abilities (Hommet et al., 2011). It is polygenic in nature and involves large number of variations in genes and their critical interactions that lead to this disease (Ray et al., 2008). Extracellular amyloid beta (Aβ) plaques, intracellular neurofibrillary tangles (NFT), cerebrovascular amyloid, dystrophic neuritis and loss of synaptic connections are standard markers of neurodegeneration in AD (Panigrahi and Singh, 2012; Tarawneh and Holtzman, 2010). It is found that the aberrant toxic Aβ aggregation causes synaptic dysfunction, oxidative stress, ionic dyshomeostasis, tau aggregation, and apoptosis (Hardy and Selkoe, 2002). The cytotoxic Aβ fibril is one of the coherent contenders for causing the starting damage to neurons in AD (Carter and Chou, 1998). Another noticeable fact is that AD does not affect entire brain at once as middle temporal gyrus (MTG) shows early AD pathology compared to the other regions of brain like entorhinal cortex (EC), hippocampus (HIP), and posterior cingulate cortex (PCC) (Ray and Zhang, 2010).

Various techniques have been used for the analysis of gene expression data associated with neurodegenerative disorders like AD. DNA microarrays constitute a contemporary tool for hypothesis generation (Newman and Weiner, 2005). Using this technique large amount of gene expression data have been easily accumulated. Now the challenging task is to extract valuable biological information from this immense amount of data (Kong et al., 2011). It is done either by identifying 'critical genes' which might singlehandedly produce a biological effect or by finding patterns in the list that point to an underlying biological process. Then annotating each gene on the list and looking for groups of genes that share a particular characteristic (Stekel and Bioinformatics, 2003). This shared or interacting nature of genes is crucial for the analysis of complex polygenic disease like AD. The development of microarray technology provides researchers, a tool that measures the expression levels of thousands of genes at once, offering possible molecular clues regarding mechanisms underlying the disease pathophysiology (Huang et al., 2009). The Gene Ontology (GO) consortium has brought systematic order to the field of gene annotation by pre-categorizing genes by biological process, molecular function, and cellular component (Ashburner et al., 2000).

The focus of bioinformatics development has now shifted from understanding networks encoded by model species to understanding the networks underlying human disease, by the increase of the human protein interaction data (Kann, 2007). Combining these network-based disease studies with the original analyses of network properties in model organisms may override the conclusion that genes associated with a particular phenotype or function, including the progression of disease, are not randomly positioned in the network. Rather, they tend to exhibit specific patterns such as high connectivity, cluster together, and occur in central network locations. There are evidences based on network property values where it has been concluded that overall degree or average distance to one another tends to lie between essential and nonessential genes, and provide patterns for the inclusion of all available interacting partners for a specific biological network (Said et al., 2004; Shachar et al., 2008). Network motifs play a central role in the identification and analysis of such specific patterns in biological networks and yield significant new insights into understanding complex biological processes involved in the intricate human disease such as AD.

Recently some studies have been proposed which have commonalities in methods, while objectives are discrete (Miller et al., 2008; Ray and Zhang, 2010). Additionally approaches as well as findings are novel in all of these studies including ours. Also there are scientific proofs and suggestions for common features between AD and other diseases to establish a link and common pathogenic mechanisms for the treatment strategies (Gotz et al., 2009). Additionally several efforts have been made recently to investigate AD by using myriad computational approaches (Chou, 2004, 2005; Wei et al., 2005; Gu et al., 2009). This overall coordination of studies designates the functional commonalities for the complex mechanisms involved in AD and its links to other diseases and suggest common prediction practices and treatment strategies. Objective of this study was to find out the relationship between one of the most threatening disease AD with the normal AG or in other words the impact of AG factor on this disease. This study also identifies genes with differential topology and their further association with other biological processes regulating causing factors for AD, aging and other diseases. This analysis has been performed by applying integrative approach on various aspects of molecular data, markers, and networks to study a complex disease AD. This analysis has implications and applications for early AD detection and novel marker identification for AD.

2. Materials and methods

2.1. Data

Three separate microarray data sets were used in this study: one consists of microarrays assessing gene expression from the CA1 region of the hippocampus from 31 individuals, comprising nine controls, seven with incipient AD, eight with moderate AD, and seven with severe AD (Blalock et al., 2004). Second data set is of 30 microarrays representing a study of the effects of aging on frontal lobe gene expression of individuals who died of natural causes between the ages of 26 and 106 (Lu et al., 2004). The AD study used Affymetrix HG-U133A chips containing 22,283 probe sets, and the aging study used HG-U95A chips with 12,625 probe sets. In addition to these data sets, one more dataset used for comparative analysis consists of 14 normal controls and 14 AD affected samples obtained from Gene Expression Omnibus (GEO Accession Number: GDS2601) (Maes et al., 2007). Additionally to incorporate network profile and network motif studies, network/pathways data associated with AD has been utilized from KEGG and other popular interaction resources.

2.2. Data pre-processing

Data filtering or normalization can reduce the dataset by removing poor or questionable data, data deemed uninteresting or irrelevant to the analysis. In this study normalization of the datasets was done using one of the tool of TM4 called Microarray Data Analysis System (MIDAS) and normalization modules used were locally weighted linear regression (Cleveland and Devlin, 1988) and total intensity normalization (Yang et al., 2002). The factors considered in the filtration of dataset include low-intensity cutoff, intensity-dependent Z-score cutoffs and replicate consistency trimming, creating a highly customizable method for preparing expression data for subsequent comparison and analysis. MIDAS provides scatter plots that illustrate the effects of each algorithm on the data (Saeed et al., 2003). Preprocessed data was subjected to individual differential gene expression followed by manually scrutinized.

2.3. Differential gene expression

Differential expression of probe sets for each dataset was performed using significance analysis of microarrays (SAM) (Tusher et al., 2001). This supervised learning software for genomic expression data mining determines differentially expressed genes in a two class