Computers in Biology and Medicine 113 (2019) 103380

Contents lists available at ScienceDirect

Computers in Biology and Medicine

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

Comparison of Methods for Differential Co-expression Analysis for Disease Biomarker Prediction

Tulika Kakati[a], Dhruba K. Bhattacharyya[a,*], Pankaj Barah[b], Jugal K. Kalita[c]

[a] Department of Computer Science and Engineering, Tezpur University, Tezpur, Assam, 784028, India
[b] Department of Molecular Biology and Biotechnology, Tezpur University, Tezpur, Assam, 784028, India
[c] Department of Computer Science, University of Colorado, Colorado Springs, CO, 80918, USA

ARTICLE INFO

Keywords:
Differential co-expression analysis
Empirical study
Gene expression
miRNA expression
Alzheimer's disease
Parkinson's disease
Disease biomarkers

ABSTRACT

In the recent past, a number of methods have been developed for analysis of biological data. Among these methods, gene co-expression networks have the ability to mine functionally related genes with similar co-expression patterns, because of which such networks have been most widely used. However, gene co-expression networks cannot identify genes, which undergo condition specific changes in their relationships with other genes. In contrast, differential co-expression analysis enables finding co-expressed genes exhibiting significant changes across disease conditions. In this paper, we present some significant outcomes of a comparative study of four co-expression network module detection techniques, namely, THD-Module Extractor, DiffCoEx, MODA, and WGCNA, which can perform differential co-expression analysis on both gene and miRNA expression data (microarray and RNA-seq) and discuss the applications to Alzheimer's disease and Parkinson's disease research. Our observations reveal that compared to other methods, THD-Module Extractor is the most effective in finding modules with higher functional relevance and biological significance.

1. Introduction

Identification of gene and miRNA biomarkers associated with diseases is an emerging and challenging task. A particular miRNA can target several genes and control their expressions and functions, thereby behaving as potent biological regulators. In other words, miRNAs regulate the expression levels of genes, potentially causing abnormal changes during progression of a disease. Therefore, identification of genes and miRNAs as disease biomarkers has impacted positively on the ability of researchers to gain substantial knowledge about how biomarkers influence the targets.

Widespread use of microarray and RNA-seq technologies has resulted in rapid growth of biological data related to gene expression and miRNA expression, thereby increasing the demand for computational methods and tools for better transcriptome analysis. These computational approaches help decipher the complex mechanisms of diseases potentially aiding the development of effective drug targets to prevent the pathogenesis of the disease at an early stage. Among these computational methods, gene co-expression network (gCEN) is widely used in functional prediction because a co-expression network (CEN) is good at identifying functionally co-related genes and miRNAs across a group of samples. Similarly, miRNA co-expression networks (m_i CEN) are used to study correlation of miRNAs with regulatory genes and identify their roles in dysregulation of expression, influencing activities, such as regulation of diseases [1,2]. However, identification of disease related miRNAs using m_i CEN has not been fully explored, and therefore assesment of methods to construct m_i CEN has become critical. In this work, we assess the performance of four CEN module extraction techniques, namely THD-Module Extractor [3], DiffCoEx [4], MODA [5], and WGCNA [6] in mapping both gene expression and miRNA expression data to CEN. The analysis of the co-expressed modules involves Gene Ontology (GO) enrichment analysis, topology enrichment, KEGG enrichment analysis, hub gene identification, and gene regulatory network analysis [7]. However, an issue in CEN analysis is that it is not able to identify the genes or miRNAs, which undergo condition-specific changes across disease conditions. A traditional approach to identify genes and miRNAs showing varying expression patterns across conditions is differential expression (DE) analysis. However, DE ignores potential interactions among these bio-entities and only considers individual entity behaviors across different conditions. To address the inter-dependence issue, an approach based on co-expression analysis, called differential co-expression (DC) analysis has emerged, referring to the condition specific changes of co-expressed modules. DC analysis provides insights into the altered mechanisms between control and

* Corresponding author.
E-mail address: dkb@tezu.ernet.in (D.K. Bhattacharyya).

https://doi.org/10.1016/j.compbiomed.2019.103380
Received 17 May 2019; Received in revised form 1 August 2019; Accepted 3 August 2019
Available online 10 August 2019
0010-4825/

disease conditions by analyzing the gene expression patterns. The main idea behind this analysis is that co-expressed or co-regulated genes or miRNAs have similar expression patterns and during progression of a disease, these co-regulated genes or miRNAs undergo mutations, which may lead to distortions in their bonds with other genes or miRNAs, thereby resulting in differences in expression patterns of the co-regulated genes across disease conditions [8].

With the continuous evolution in high throughput technologies, such as, microarray and RNA-sequencing, there is a tremendous growth in the amount of biological data produced. Microarray data record expression values of thousands of genes simultaneously and their analysis helps in getting insights into the changes in expression values of multiple genes during the progression of a disease. The genes whose expression values changes exhibit altered biological patterns across stages of a disease and therefore act as biomarkers. Gene biomarkers are groups of genes which exhibit distinct features during progression of the disease. The expression of the gene biomarkers may upregulate or down-regulate across stages of diseases. On the other hand, in next generation sequencing technologies (NGS), a large amount of DNA sequencing ican be performed at a cheaper cost and quantification of discrete read count is aligned to a particular sequence. Co-expression analysis of NGS (RNA-seq) data provides better structural variations. According to Dam et al. [7], RNA-seq based CEN module extraction techniques share a number of limitations including loss of information, biased co-expression signals, and increase in the size of networks. Conesa et al. [9] review the steps of RNA-seq data analysis and address the challenges and issues including quantification, read count, visualization, differential expression, and functional analysis. Despite major challenges, a number of CEN module extraction techniques have been introduced using RNA-seq data. Among them, WGCNA is the pioneer in detecting co-expressed modules from RNA-seq data using CEN. Kogelman et al. [10] identify co-expressed modules of genes using WGCNA in porcine RNA-seq data and confirm the complexity in obesity by identifying complex networks and pathways. Another method proposed by Lee et al. [11] is used to analyze RNA-Seq data and to detect and quantify individual exons and transcripts. Wang and Liu [12], report a machine learning model that identifies differentially co-expressed modules from a CEN built from gene expression profiles of different brain regions in order to mine Alzheimer's disease (AD) biomarkers having perturbations across control and AD conditions.

CEN techniques for microarray and RNA-Seq data have both differences and similarities. There are a few recent surveys on commonly used CEN and DC analysis techniques. Van Dam et al. provided an introduction to CEN, followed by a comprehensive study of the commonly used CEN tools and the scope of DC analysis using both microarray and RNA-seq data [7]. Hussain et al. [13] present a comprehensive survey by reviewing differential co-expression, co-expression, differential networking, differential connectivity, and scRNA-data analysis for both microarray and RNA-seq data. Lie et al. [14] review existing methods for DC analysis and discuss their application to cancer datasets. However, there is a need for an empirical study of the most commonly used DC analysis techniques with respect to their significance in producing biologically and statistically meaningful results. In this review, we provide an overview of four module extraction techniques, namely THD-Module Extractor, DiffCoEx, MODA, and WGCNA for both microarray and RNA-Seq data using gene and miRNA expression profiles (see Fig. 1) and discuss their application in identifying gene and miRNA biomarkers related to Alzheimer's disease (AD) and Parkinson's disease (PD).

We present some significant outcomes of an empirical study of the four aforementioned techniques for identifying differentially co-expressed genes and miRNAs, and compare results in terms of GO enrichment, topological enrichment (for differentially co-expressed genes), miRNA enrichment (for differentially co-expressed miRNAs), and pathway enrichment. In addition, from each differentially co-expressed gene or miRNA modules, we extract the hub genes and validate their association with the disease. DC analysis techniques identify co-expressed genes, which undergo dysfunctionality during transition from control to disease conditions. We use the four DC analysis techniques on three microarray and four RNA-seq datasets with both gene and miRNA expression profiles. Further, we extend DC analysis to identify hub genes and validate their dysfunctional characteristics across control and disease conditions and their associations with AD and PD.

2. Methods

There are a number of gCEN module extraction techniques to infer gene functions and identify disease-gene associations from gene expression data. Due to short lengths of miRNAs and high variance and small sizes of the miRNA datasets, there are few m_i CEN module extraction techniques. In our study, we use four module extraction techniques, namely THD-Module Extractor, DiffCoEx, MODA, and WGCNA to extract both gene and miRNA co-expressed modules due to their varied underlying approaches and attributes. Among these techniques, THD-Module Extractor and WGCNA do not assume the presence of abnormality between control and disease conditions. To address this issue, we extend these traditional CEN techniques to identify variations among genes and miRNAs under control and disease conditions using differential co-expression functions along with DiffCoEx and MODA.

2.1. THD-Module Extractor

THD-Module Extractor is a CEN module extraction method which considers the border genes (genes with single edges) along with other co-expressed genes in disease related analysis. The method mines interesting border genes from co-expressed modules, which have high semantic correlation and low expression similarity with the core gene, which is the gene with the highest number of edges. The method can be applied to microarray AD datasets to identify disease biomarkers. The method accepts two thresholds, viz., expression similarity threshold (δ) and minimum neighborhood threshold (ρ), and constructs CENs from gene expression data [3]. In each iteration of module extraction, δ is updated by a factor α to detect modules from the CEN. Unlike WGCNA, the user can tune the minimum number of genes in a co-expressed module and the number of modules generated. The method has been implemented in both MATLAB and R. THD-Module Extractor has been extended to find the differentially co-expressed genes and miRNAs by finding covariance among genes and miRNAs across disease conditions.

2.2. DiffCoEx

DiffCoEx builds CENs based on WGCNA and follows five steps to identify differentially co-expressed genes. At first, for each condition, it constructs an adjacency matrix using Pearson correlation and then computes the matrix of adjacency differences for pairs of conditions. From the matrix of adjacency differences, we compute a dissimilarity matrix from topological overlap scores, where high scores signify genes with same neighbors. From the dissimilarity matrix, co-expressed modules are found. The method depends on the tuning of a soft threshold parameter used to construct adjacency matrix and tree cutting thresholds [4]. The method is implemented in R and is used as an application for microarray datasets only.

2.3. MODA

MOdule Differential Analysis for weighted gene co-expression network (MODA) is a CEN module extraction technique based on WGCNA to identify differentially expressed modules of genes, which are associated with a disease. MODA uses a sample-saving approach to construct CEN from single or multiple samples and identify the differentially co-expressed gene modules across conditions by comparing networks across conditions [5]. In order to detect modules from each

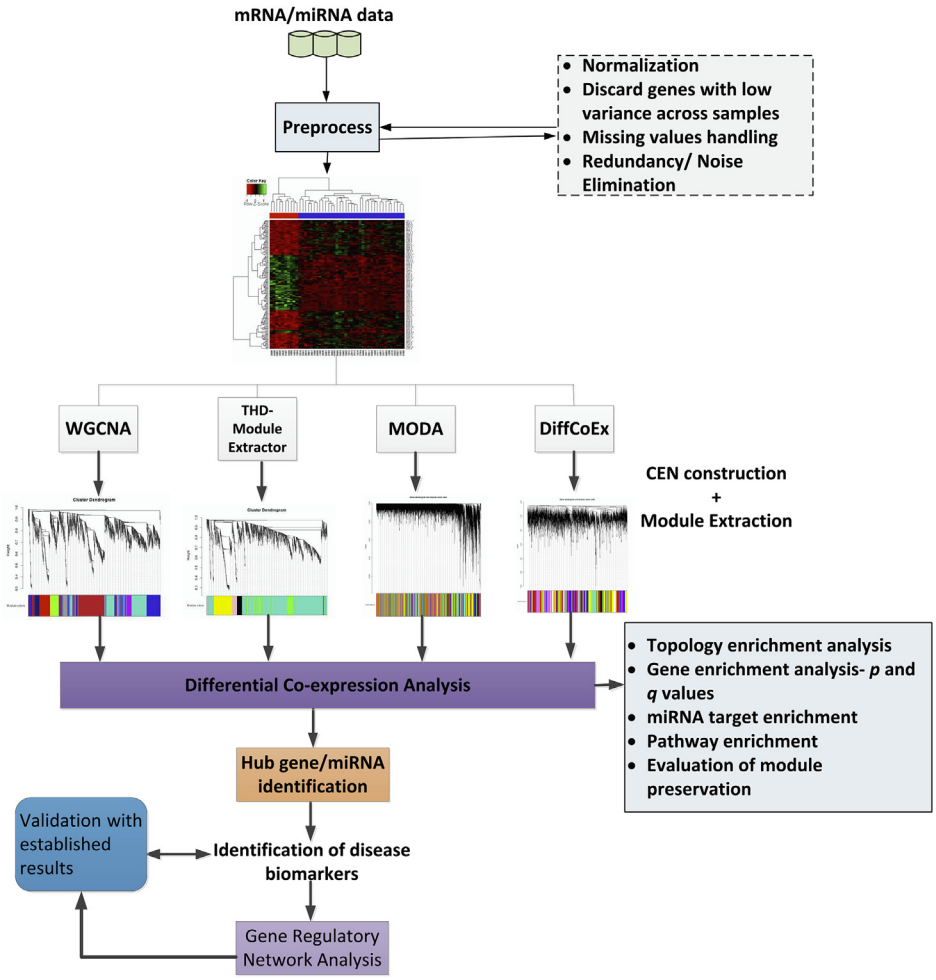

Fig. 1. Framework shows the steps in identifying gene or miRNA biomarkers using the four module extraction techniques following DC analysis. The data are pre-processed using normalization, mean, variance, handling missing or NAN values, removing noise or redundant entities, before proceeding with network construction and module extraction. The lowly expressed or non-varying genes and miRNAs usually represent noise. For ease of computation, we filter genes and miRNAs with low variance with respect to a threshold. After extraction of gene or miRNA co-expressed modules, we extend co-expression analysis with differential co-expression (DC) analysis. The differentially co-expressed genes or miRNAs are further analyzed for topology/miRNA enrichment analysis, GO enrichment analysis, and pathway enrichment. Further, from the differentially co-expressed genes or miRNAs, hub genes or miRNAs are identified and their associations with a particular disease are validated with already published results. We also find regulatory behavior of these hub genes or hub miRNAs in disease networks. These regulatory hub genes or hub miRNAs are found to play significant roles in disease progression.

network, hierarchical clustering is applied by tuning the tree cutting threshold. The optimal cutting threshold is decided based on average modularity for weighted networks. The method is implemented in R and the results are found to be highly influenced by parameter values.

2.4. WGCNA

WGCNA or Weighted gene co-expression network analysis uses Pearson correlation to compute correlation patterns among genes across all samples or conditions. WGCNA uses hierarchical clustering to identify modules from the CEN using different tree-cutting thresholds by summarizing each module by the module eigengene or an intramodular hub gene [6]. The eigengene helps relate modules by computing memberships of modules. WGCNA is the most widely used CEN technique and is used in many applications, such as, genetic analysis of cancer, genome analysis of mouse, and yeast, and analysis of brain MRI data. Kogelman et al. introduced a CEN model to identify genes or pathways such as NKT (Natural Killer cell mediated cytotoxity), osteoporosis, and B-cell receptor signaling with low p values [10]. The method was implemented in R and the underlying methods can be used in both microarray gene expression data and RNA-Seq data. To find differentially co-expressed genes or miRNAs, we use *WGCNA()* function from the DGCL package.

Table 1 highlights the main features of the four CEN module extraction techniques. DiffCoEx and MODA are based on the WGCNA framework and they suffer from the major disadvantages of WGCNA, listed below.

- WGCNA does not recommend that one produce modules from a dataset consisting of fewer than 15 samples. THD-Module Extractor does not place a limit on datasets with a small number of samples.
- WGCNA does not assume scale-free topology for differentially expressed genes and therefore the soft thresholding of power fit fails. On the other hand, for THD-Module Extractor, there is no such assumption for differentially expressed genes.
- WGCNA takes many parameters with default values to control network construction and distinct module extraction. Tuning of these parameters is situation specific. THD-Module extractor takes only three parameters, namely expression similarity threshold (δ), minimum neighborhood threshold (ρ), and δ updating factor α for network construction and module extraction. The range for δ and ρ is [0.95–0.35] and default value of α is 0.05.

In our study, we use seven datasets (four RNA-seq datasets and three microarray datasets) with both gene and miRNA expression profiles for (differential) co-expression analysis of the modules. We extend the analysis of the differentially co-expressed genes with GO enrichment, topology enrichment (for differentially co-expressed genes), miRNA target enrichment (for differentially co-expressed miRNAs), pathway enrichment, and identification of hub genes. The hub genes extracted from the differentially co-expressed gene or miRNA modules undergo regulatory changes across the disease conditions. We identify the hub genes and hub miRNAs, which undergo disorientation in the disease network in terms of network statistics, such as the degrees of the hub genes. We further assess the condition specific hub genes and hub miRNAs and identify their potential roles in the progression of AD and PD. We also use network preservation statistics, $Z_{summary}$ to report the evidence for weak module preservation of the differentially co-

Table 1
Comparison of four CEN module extraction techniques.

Method	Working principle	Measure used in CEN construction	Parameters	Implementation language or Platform used	Datasets used
THD-Module Extractor [3]	Density-based clustering	SSSim [15]	Expression similarity threshold (δ), minimum neighborhood threshold (ρ), and δ updating factor α	*R* and MATLAB	Microarray
DiffCoEx [4]	Differential Co-expression analysis	Dissimilarity measure based on TOM [16]	Soft threshold parameter and tree cutting threshold	*R*	Microarray
MODA [5]	Sample-saving approach	Pearson correlation or custom distance measure (same as WGCNA)	Tree cutting threshold	*R*	Microarray
WGCNA [6]	Hierarchical clustering	Pearson correlation or custom distance measure	Tree-cutting thresholds	*R*	Microarray and RNA-Seq

expressed gene and miRNA modules.

Table 2 gives an overview of the microarray and RNA-seq datasets used in our experimental study. For microarray datasets, we used two gene expression AD datasets (GSE4757, GSE5281) and one miRNA expression AD dataset (GSE16759). For RNA-Seq datasets, we used two gene expression PD datasets (GSE62642, GSE68719) and two miRNA expression PD datasets (GSE72962, GSE77668).

WGCNA finds good genes or miRNAs using *goodSamplesGenes()*. WGCNA does not recommend that we use differential expression analysis before module extraction. A major drawback of high throughput technologies such as microarray and RNA-seq is the batch effect, which is incorporated due to differences in dates of experiments, instruments used, populations, and technicians. In microarray data, batch effect comes from unbalanced experiments, while in RNA-Seq batch effects come from unbalanced experiments and sequence design. The variance in RNA-seq is bigger than in microarray data, which makes the batch effects in RNA-seq more prominent. In our experiments, we remove batch effects for RNA-seq data using *removeBatchEffect()* of the edgeR package. The preprocessing activity of microarray and RNA-seq datasets involves the following steps.

- We normalize the logarithmic expression data using *normalize.quantiles()*
- We remove redundant genes and miRNAs. We also filter out the genes and miRNAs which have NAN values.
- For computational ease, we scale down the number of genes and miRNAs by using a naive approach called *varFilter()* and discard the genes and miRNAs which had very low variance.

3. Results

In this section, we report the assessment of the four module extraction techniques in terms of GO enrichment, topology enrichment, miRNA target enrichment and pathway enrichment. Further, we extended the analysis with identification of hub genes and hub miRNAs across control and disease networks of GSE68719 and GSE72962 PD datasets, respectively. We assess the gene regulatory network of the hub genes, identified from differentially co-expressed genes obtained from GSE4757 AD dataset. We also report the module preservation statistics of the differentially co-expressed gene modules across control and disease networks of GSE4757 AD dataset.

3.1. Differentially co-expressed gene modules

We study the Gene Ontology (GO) and the topological significance of the co-expression modules extracted from both AD and PD datasets. Statistical measures, such as p and q values, define the potential GO enrichment of a module. The p value is defined as the possibility of obtaining k or more genes in a cluster of size n, associated with a specific GO term. q value is another statistical measure, which gives the minimum false discovery rate (FDR). Lower values of p and q signify higher statistical significance for a module, i.e., that the module was not discovered by chance. The topological measures, such as shortest path length (SPL), node betweeness (NB), and degree determine the topological enrichment for the modules. SPL measures the centrality of a node with respect to other nodes passing through that node. Lower SPL means higher interaction between the nodes of a module. The NB of a module measures the number of shortest paths passing through a node. The higher the value of NB, the more significant is the presence of the node in the module. The degree of a module can be used to compare a node with another in terms of the average number of edges incident on it. A higher degree signifies a higher quantity of interactions among genes of a module. Pathway analysis measures the importance of genes or groups of genes which undergo differential expression during pathogenesis of a disease. We show the significance of the KEGG pathways associated with the differentially co-expressed gene modules extracted

Table 2
Overview of the datasets used in our experimental study https://www.ncbi.nlm.nih.gov/geo/.

Technology	Disease	ID: gene/miRNA	original dataset size	preprocessed dataset size
Microarray	AD	GSE4757: gene-expression	54675 genes × 10 control conditions × 10 disease conditions	3511 × 10 control conditions × 10 disease conditions
		GSE5281:gene-expression	54675 genes × 74 control conditions × 87 disease conditions	5070 × 74 control conditions × 74 disease conditions
		GSE16759:miRNA expression	899 genes× 4 control conditions × 4 disease conditions	82 × 4 control conditions × 4 disease conditions
RNA-Seq	PD	GSE62642: gene-expression	25593 genes × 8 control conditions × 6 disease conditions	9080 × 6 control conditions × 6 disease conditions
		GSE68719: gene-expression	17580 genes × 44 control conditions × 29 disease conditions	9080 × 29 control conditions ×29 disease conditions
		GSE72962:miRNA-expression	907 genes × 33 control conditions × 29 disease conditions	907 × 29 control conditions ×29 disease conditions
		GSE77668:miRNA-expression	822 genes × 12 control conditions × 12 disease conditions	518 × 12 control conditions ×12 disease conditions

using the four module extraction techniques in terms of *p* and *q* values.

3.1.1. GO and topology enrichment analysis

3.1.1.1. AD datasets

i. *Performance of THD-Module Extractor:* For GSE4757 AD dataset, the *p* and *q* values obtained for the best differentially co-expressed gene module extracted using THD-module Extractor are 1.96E-06 and 1.38E-03, respectively. This module is also topologically enriched with SPL, NB, and degree values of 3.97, 20666, and 11.19, respectively. For GSE5281 AD dataset, THD-Module Extractor also performs well in extracting differentially co-expressed gene module with lower *p* and *q* values of 7.72E-07 and 9.50E-03, respectively and SPL, NB, and degree values of 4.21, 11895, and 8.03, respectively.

ii. *Performance of DiffCoEx:* For GSE4757 AD dataset, the best differentially co-expressed gene module extracted using DiffCoEx has *p* and *q* values of 2.79E-05 and 2.75E-02, respectively. The SPL, NB, and degree values of this module are 4.03, 20585, and 10.79, respectively. Similarly, for GSE5281 AD dataset, the *p*, *q*, SPL, NB, and degree values of the best differentially co-expressed module extracted using DiffCoEx are 5.10E-06, 3.06E-02, 4.06, 21094, and 10.58, respectively.

iii. *Performance of MODA:* The best differentially co-expressed gene module obtained from the GSE4757 AD dataset using MODA has *p*, *q*, SPL, NB, and degree values of 6.85E-06, 2.45E-03, 3.99, 44151, and 21, respectively. For GSE5281 AD dataset, the *p*, *q*, SPL, NB, and degree values of the best differentially co-expressed module extracted using MODA are 1.03E-04, 2.85E-02, 3.92, 10897, and 9.03, respectively.

iv. *Performance of WGCNA:* The best differentially co-expressed gene module extracted from GSE5281 AD dataset using WGCNA has *p*, *q*, SPL, NB, and degree values of 6.71E-05, 1.36E-02, 4.19, 2425, and 4.94, respectively. Similarly, for GSE5281 AD dataset, the differentially co-expressed gene module extracted using WGCNA has *p*, *q*, SPL, NB, and degree values of 7.10E-06, 2.24E-03, 4.03, 9670, 8.41, respectively.

v. *Comparison:* In Figs. 2 and 3, we compare the statistical and topological enrichment of the differentially co-expressed gene modules, extracted from the GSE4757 and GSE5281 AD datasets, respectively using the four techniques, in terms of *p* value, *q* value, SPL, NB, and degree for the modules. In our experimental study for GSE4757, we find that the best differentially co-expressed module in terms of higher statistical significance and topological enrichment is obtained by THD-Module Extractor. The topological values for differentially co-expressed module from MODA are also significant. Similarly, for GSE5281 dataset, we observe that the best differentially co-expressed gene module in terms of statistical significance is extracted using THD-Module Extractor. However, in terms of topological enrichment, the best differentially co-expressed module is extracted by DiffCoEx. Again, for MODA except SPL, other topological statistics, viz., NB, and degree and GO enrichment parameters, viz., *p* and *q* values are not that significant. For WGCNA, the best differentially co-expressed modules is found to be rich with GO terms, namely *cellular nitrogen compound metabolic process, ion binding, and gene expression* with lower *p* and *q* values of 7.10E-06, 2.24E-03, respectively. Studies show that GO terms such as cellular nitrogen compound metabolic process, ion binding, gene expression with lower *p* values are associated with AD. Chatterjee et al. [17] established that AD-related genes such as STXBP1 and GASP are enriched in GO terms such as gene expression. From a molecular perspective, there is a significant interference of ions during progression of AD and the perturbation of ions is related to GO term ion binding [18].

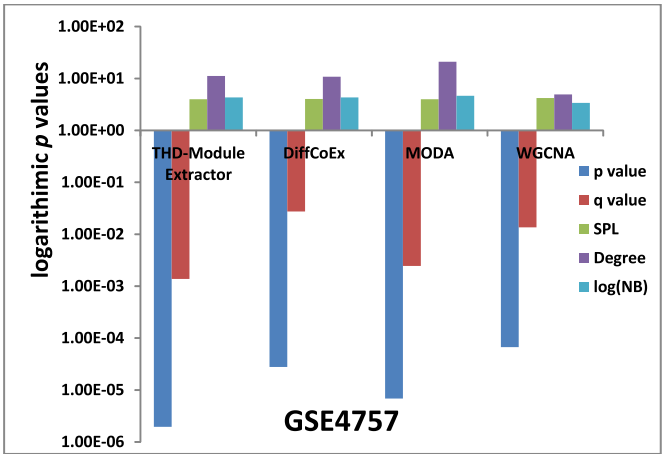

Fig. 2. Comparison of statistical and topological enrichment for the best modules extracted using THD-Module Extractor, DiffCoEx, MODA, and WGCNA for GSE4757. The y-axis denotes the logarithmic values of *p*, *q*, SPL, NB, and degree and the x-axis denotes the four methods, namely THD-Module Extractor, DiffCoEx, MODA, and WGCNA.

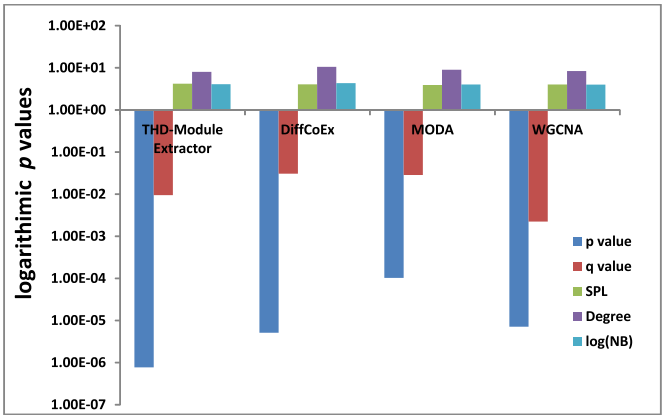

Fig. 3. Comparison of statistical and topological enrichment for the best modules extracted using THD-Module Extractor, DiffCoEx, MODA, and WGCNA for GSE5281. The y-axis denotes the logarithmic values of *p*, *q*, SPL, NB, and degree and the x-axis denotes the four methods, namely THD-Module Extractor, DiffCoEx, MODA, and WGCNA.

3.1.1.2. PD datasets

i. *Performance of THD-Module Extractor:* The *p* and *q* values for the best differentially co-expressed gene module extracted from GSE62642 PD dataset using THD-module Extractor are 1.24E-05 and 7.30E-03, respectively. This module is also topologically enriched with SPL, NB, and degree values of 3.97, 18381, and 11.13, respectively. Similarly, for GSE68719 PD dataset, the *p*, *q*, SPL, NB, and degree values for the best differentially co-expressed module extracted using THD-module Extractor are 2.49E-07, 5.51E-04, 4.09, 21465, and 9.61, respectively.

ii. *Performance of DiffCoEx:* For GSE62642 PD dataset, the best differentially co-expressed module extracted using DiffCoEx has *p*, *q*, SPL, NB, and degree values of 2.24E-07, 2.93E-03, 4.05, 24376, and 12.42, respectively. Similarly, the *p*, *q*, SPL, NB, and degree values of the best differentially co-expressed module for DiffCoEx are 5.52E-05, 4.39E-02, 4.02, 13765, and 8.6, respectively.

iii. *Performance of MODA:* The best differentially co-expressed gene module extracted from GSE62642 PD dataset using MODA has *p*, *q*, SPL, NB, and degree values of 7.70E-05, 7.56E-03, 4.26, 50394, and 18.33, respectively. Similarly, for GSE68719 PD dataset, the *p*, *q*, SPL, NB, and degree values of the best differentially co-expressed gene module extracted using MODA are 3.95E-06, 2.11E-03, 4.01, 20889, and 12.64, respectively.

iv. *Performance of WGCNA:* We do not find any differentially co-expressed gene module from GSE62642 PD dataset using WGCNA due to its smaller sample size. For GSE68719, the best differentially co-expressed gene module extracted using WGCNA has *p*, *q*, SPL, NB, and degree values of 4.58E-08, 5.79E-05, 4.09, 24699, and 10.92, respectively.

v. *Comparison:* In Figs. 4 and 5, we observe that for GSE62642, DiffCoEx outperforms other three methods in terms of higher statistical significance with lower *p* and *q* values of 2.24E-07 and 2.93E-03, respectively. The differentially co-expressed gene module from MODA is highly enriched in terms topological statistics, viz., a lower SPL value of 4.26, a higher NB value of 50394, and a higher degree of 18.33. For WGCNA, we do not find statistically significant and topologically enriched differentially co-expressed gene modules. On the other hand, the best differentially co-expressed gene module for GSE68719 dataset, in terms of statistical significance are extracted using WGCNA. This differentially co-expressed module is found to be rich with GO terms with lower *p* and *q* values of 4.58E-08 and 5.79E-05, respectively. The differentially co-expressed module extracted using MODA from GSE68719 has best values for topological statistics viz., SPL, NB, and degree in comparison to the differentially co-expressed modules, extracted using other three methods.

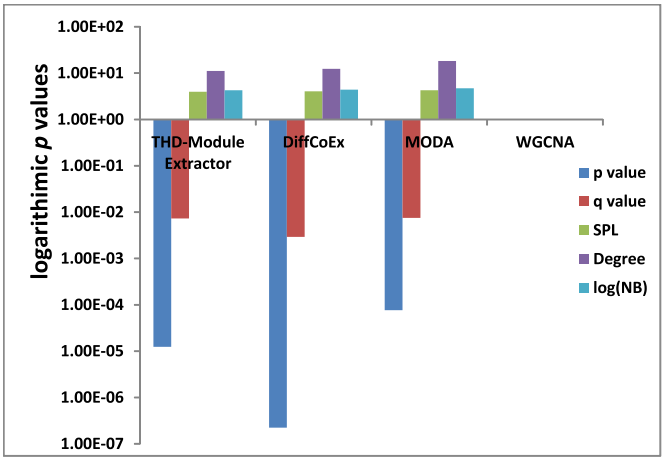

Fig. 4. Comparison of statistical and topological enrichment for the best modules extracted using THD-Module Extractor, DiffCoEx, MODA, and WGCNA for GSE62642. The y-axis denotes the logarithmic values of *p*, *q*, SPL, NB, and degree and the x-axis denotes the four methods, namely THD-Module Extractor, DiffCoEx, MODA, and WGCNA.

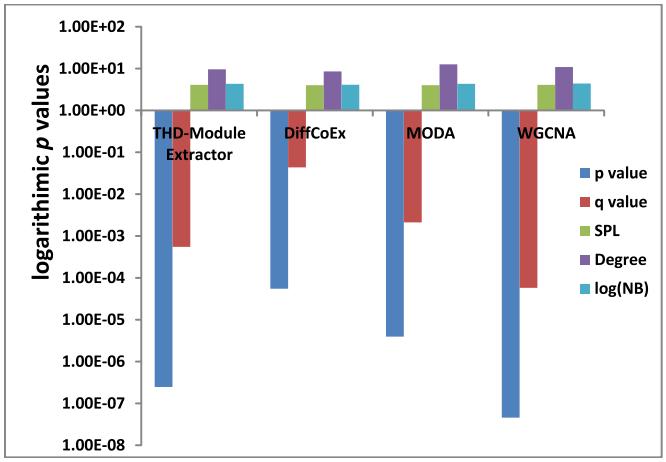

Fig. 5. Comparison of statistical and topological enrichment for the best modules extracted using THD-Module Extractor, DiffCoEx, MODA, and WGCNA for GSE68719. The y-axis denotes the logarithmic values of *p*, *q*, SPL, NB, and degree and the x-axis denotes the four methods, namely THD-Module Extractor, DiffCoEx, MODA, and WGCNA.

In Table 3, we conclude the best differentially co-expressed gene module extracted using THD-Module Extractor, DiffCoEx, MODA, and WGCNA based the above GO and topological enrichment analysis.

3.1.2. Pathway enrichment analysis

3.1.2.1. AD datasets

i. *Performance of THD-Module Extractor:* For GSE4757 AD dataset, the differentially co-expressed gene modules obtained using THD-Module Extractor are mapped to AD related pathways, such as *Alzheimer disease-amyloid secretase pathway, Wnt signaling pathway, Alzheimer disease-presenilin pathway* with better statistical significance (lower *p* and *q* values). Similarly, for GSE5281 AD dataset, the differentially co-expressed gene module extracted using this method mapped to pathways such as *Inflammation mediated by chemokine and cytokine signaling pathway, Alzheimer disease-amyloid secretase pathway and p53 pathway* related to AD.

ii. *DiffCoEx:* For both GSE4757 and GSE5281 AD datasets, DiffCoEx mines differentially co-expressed gene modules, which are mapped to AD associated pathways such as *Integrin signalling pathway, Inflammation mediated by chemokine and cytokine signaling pathway, Dopamine receptor mediated signaling pathway, Notch signaling pathway, and p38 MAPK pathway* having higher statistical significance.

iii. *MODA:* The pathways such as *Cadherin signaling pathway, Dopamine receptor mediated signaling pathway, Wnt signaling pathway, Alzheimer disease-presenilin pathway, Angiogenesis, and Notch signaling pathway* mapped to the differentially co-expressed gene modules extracted from GSE4757 and GSE5821 AD datasets using MODA are also significant in terms of their association with AD.

iv. *WGCNA:* The differentially co-expressed gene modules extracted from GSE4757 and GSE5821 AD datasets using WGCNA are also mapped to AD related pathways such as *p53 pathway, Integrin signalling pathway, Wnt signaling pathway, PDGF signaling pathway, and CCKR signaling map.*

v. *Comparison:* It is unlikely to compare the pathways mapped to differentially co-expressed gene modules extracted from each method as the pathways are not common. However, in Table 4 and table 4 of

Table 3
Performance of methods based on GO and topological enrichment.

Method	GO Enrichment				Topological Enrichment			
	GSE4757	GSE5281	GSE62642	GSE68719	GSE4757	GSE5281	GSE62642	GSE68719
THD-Module Extractor	best	best						
DiffCoEx			best			best		
MODA					best		best	best
WGCNA				best				

Supplementary file, we show that the differentially co-expressed gene modules extracted from GSE4757 and GSE5281 AD datasets, respectively using the four methods are mapped to diverse signaling pathways linked to AD, such as *Wnt signaling, p53 pathway, Alzheimer disease secretase pathway, Notch, and Inflammation mediated by chemokine and cytokine signaling pathway* [19–23].

3.1.2.2. PD datasets

i. *Performance of THD-Module Extractor:* The differentially co-expressed gene modules obtained from GSE62642 PD dataset using THD-Module Extractor are mapped to significant PD related pathways, such as *Dopamine receptor mediated signaling pathway, Inflammation mediated by chemokine and cytokine signaling pathway, and Toll receptor signaling pathway* with lower p and q values. Similarly the pathways, such as *Dopamine receptor mediated signaling pathway, Notch signaling pathway, and p38 MAPK pathway* mapped to the differentially co-expressed gene modules extracted GSE68719 PD dataset using THD-Module Extractor are significant in terms of p and q values.
ii. *Performance of DiffCoEx:* The pathways such as, *Dopamine receptor mediated signaling pathway, Notch signaling pathway, p38 MAPK pathway, Angiogenesis, Alzheimer disease-presenilin pathway* mapped to the differentially co-expressed gene modules extracted from GSE62642 and GSE68719 using DiffCoEx are PD related and statistically significant.
iii. *Performance of MODA:* The differentially co-expressed gene modules obtained from GSE62642 and GSE68719 using MODA are mapped to pathways such as *Alzheimer disease-presenilin pathway, Angiogenesis, Notch signaling pathway, Wnt signaling pathway, and Parkinson disease*, which are both AD related and statistically significant.
iv. *Performance of WGCNA:* The pathways such as *PDGF signaling pathway, CCKR signaling map, Dopamine receptor mediated signaling pathway, Wnt signaling pathway, and Parkinson disease* are mapped to the differentially co-expressed gene modules obtained from GSE62642 and GSE68719 using WGCNA with lower p and q values.
v. *Comparison:* Tables 5 and 6 of Supplementary file show the pathway enrichment of the differentially co-expressed genes modules extracted from GSE62642 and GSE68719 PD datasets using the four module extraction techniques. The pathways mapped to the differentially co-expressed gene modules obtained from these datasets are not common and therefore it is not recommended to compare the statistical significance of these mapped pathways.

3.2. Differentially co-expressed miRNA modules

In this section, we carry out enrichment analysis using Observed/Expected (O/E) ratio. The O/E ratio shows target enrichment for the differentially co-expressed miRNA modules. For a given gene X, O/E ratio [24] is calculated using Equation (1),

$$O/E_X = \frac{A}{B} \tag{1}$$

where A is the proportion of the queried miRNA(s) which are predicted to target gene X, and B is the proportion of all miRNAs in the miRSystem database [24] predicted to target gene X. The cut-off O/E ratio is set as 2 and we define three ranges of O/E [2–40], and [41-above] as low, high, and highest respectively.

3.2.1. GO enrichment and miRNA target enrichment analysis
3.2.1.1. AD datasets

i. *Performance of THD-Module Extractor:* The differentially co-expressed miRNA modules obtained from GSE16759 AD dataset using THD-Module Extractor have lower p values of 2.63E-198, 2.41E-102, and 9.55E-205 enriched with gene targets. The target genes of the differentially co-expressed AD miRNA modules for THD-Module Extractor have O/E ratio values of 18.6842, 47.3334, and 54.6154, respectively which can be attributed to the significance of the over-representation of the target genes associated with different biological activities [25–31]. Kong et al. showed that with deterioration of AD, there is change in the regulation of RPS4Y1 which co-regulates the expression of other proteins related to AD [25]. Similarly, PLA2G3 plays an important role in β-amyloid-induced developments, such as behavioral impairments and memory declination during pathogenesis of AD [28].
ii. *Performance of DiffCoEx:* The differentially co-expressed miRNA

Table 4
Pathway enrichment of differentially co-expressed genes.

Dataset/Disease	Method	Pathways	p value	q value
GSE4757 - AD	THD-Module Extractor	Alzheimer disease-amyloid secretase pathway	1.65E-04	2.36E-03
		Wnt signaling pathway	3.21E-05	1.12E-02
		Alzheimer disease-presenilin pathway	2.11E-03	1.14E-02
	DiffCoEx	Integrin signalling pathway	7.15E-04	4.32E-02
		Inflammation mediated by chemokine and cytokine signaling pathway	2.13E-03	2.47E-02
		Notch signaling pathway	3.52E-04	1.02E-02
	MODA	Cadherin signaling pathway	6.15E-04	5.15E-02
		Dopamine receptor mediated signaling pathway	4.15E-04	3.25E-03
		Wnt signaling pathway	5.75E-04	2.15E-02
	WGCNA	p53 pathway	1.02E-05	2.78E-04
		Integrin signalling pathway	2.58E-04	1.78E-03
		Wnt signaling pathway	3.85E-04	7.85E-03

module extracted from GSE16759 AD dataset using DiffCoEx has *p* value of 3.86E-10 with O/E ratio of 26.2963. For DiffCoEx, the target genes of the differentially co-expressed AD miRNA modules, such as A2BAR, HTR1B, and NRTN play significant roles in biological processes including cell differentiation, and neuronal survival, which are important during the pathogenesis of AD [32].

iii. *Performance of MODA:* The differentially co-expressed miRNA module extracted from GSE16759 AD dataset using MODA has *p* value of 1.83E-27 with O/E ratio of 32.2727. The broad involvement of genes such as A2BAR, HTR1B, KRT126P and PDIA5 targeted by the differentially co-expressed AD miRNA modules extracted using MODA in human neurodegenerative diseases provides insight into their roles in AD intervention.

iv. *Performance of WGCNA:* The differentially co-expressed miRNA module extracted from GSE16759 AD dataset using WGCNA has *p* value of 4.70E-12 with O/E ratio of 2.8629. The target genes associated with differentially co-expressed AD miRNA modules extracted using WGCNA such as A2BAR, HTR1B, and SLC5A11 have positive correlations with the pathogenesis of AD. For example, A2BAR is involved in several biological functions such as immune response, inflammatory conditions, and cell growth in AD brains [33]. The target genes such as SLC5A11 and HTR1B correlate with many immune-related genes, which is suggestive of its role in AD pathogenesis [34,35].

v. *Comparison:* In Table 5, we compare the GO enrichment and miRNA target enrichment of the differentially co-expressed miRNA modules, extracted from GSE16759 AD, GSE72962 PD and GSE77668 PD using the four techniques in terms of *p* value and O/E ratio.

For all four methods, it is seen that these differentially co-expressed miRNA modules extracted from AD dataset are rich in target genes associated with AD with lower *p* values. However, the biological significance of the differentially co-expressed modules in terms of target enrichment (O/E) for DiffCoEx, MODA, and WGCNA is not satisfactory. This is because the sample size of GSE16759 AD is small for biologically significant module extraction. For THD-Module Extractor, the best differentially co-expressed module mapped from GSE16759 AD miRNA dataset has *p* value and O/E ratio as 9.55E-205 and 54.6154 (highest), respectively. Other differentially co-expressed modules do not have significant statistical and O/E ratios in comparison to that of THD-Module Extractor.

3.2.1.2. PD datasets

i. *Performance of THD-Module Extractor:* The best differentially co-expressed miRNA modules obtained from GSE77668 and GSE72962 PD datasets using THD-Module Extractor have *p* values of 2.94E-111 and 2.94E-27 with O/E ratios of 22.5396 and 4.55128, respectively.

ii. *Performance of DiffCoEx:* The best differentially co-expressed miRNA modules extracted from GSE77668 and GSE72962 PD datasets using DiffCoEx have *p* values of 3.11E-11 and 2.94E-111 with O/E ratios of 14.4898 and 7.39584, respectively.

iii. *Performance of MODA:* The best differentially co-expressed miRNA modules extracted from GSE77668 and GSE72962 PD datasets using MODA have *p* values of 2.94E-88 and 3.86E-10 with O/E ratio of 6.12069 and 3.81721, respectively.

iv. *Performance of WGCNA:* The best differentially co-expressed miRNA modules extracted from GSE77668 and GSE72962 PD datasets using WGCNA have *p* values of 2.94E-111 and 1.16E-06 with O/E ratio of 3.73684 and 7.63441, respectively.

v. *Comparison:* In Table 5, it is seen that for GSE77668 PD miRNA dataset, both THD-Module Extractor and WGCNA have statistically significant module with 2.94E-111 as *p* value. However, the O/E ratio for this statistically significant module extracted using WGCNA is lower than that of THD-Module Extractor. Therefore, the differentially co-expressed modules extracted from WGCNA, DiffCoEx, and MODA have lower biological significance than those of THD-Module Extractor. Again for GSE72962 PD miRNA dataset, DiffCoEx produced a module with the lowest *p* value of 2.94E-111. However, O/E ratio of this differentially co-expressed module is lower than that of THD-Module Extractor.

Table 5
Statistical and miRNA target enrichment of differentially co-expressed miRNAs.

Dataset/Disease	Method	*p* value	Target Enrichment score (O/E)
GSE16759-AD	THD-Module	2.63E-198	low
	Extractor	2.41E-102	highest
		9.55E-205	highest
	DiffCoEx	3.86E-10	high
	MODA	1.83E-27	high
	WGCNA	2.07E-05	low
GSE77668-PD	THD-Module	2.94E-111	high
	Extractor	8.73E-86	low
		3.32E-54	low
	DiffCoEx	3.11E-11	low
	MODA	2.94E-88	low
	WGCNA	2.94E-111	low
GSE72962-PD	THD-Module	3.86E-10	high
	Extractor	2.94E-27	low
	DiffCoEx	2.94E-111	low
	MODA	3.86E-10	low
	WGCNA	1.61E-06	low

3.2.2. Pathway enrichment analysis

3.2.2.1. AD datasets

i. *Performance of THD-Module Extractor:* Pathways such as Wnt Signaling and MAPK with *p* values 6.00E-17 and 1.92E-06, respectively are mapped from differentially co-expressed miRNA modules extracted from GSE16759 AD dataset using THD-Module Extractor. These pathways are related to AD.

ii. *Performance of DiffCoEx:* The differentially co-expressed miRNAs extracted from GSE16759 AD dataset using DiffCoEx are mapped to pathways such as Fatty acid metabolism and ECM-receptor interaction with *p* values 1.88E-16 and 0.001548665, respectively.

iii. *Performance of MODA:* The differentially co-expressed miRNAs extracted from GSE16759 AD dataset using MODA are mapped to pathway namely Fatty acid biosynthesis, with *p* value 1.83E-27.

iv. *Performance of WGCNA:* Pathways such as Glycosphingolipid biosynthesis - lacto and neolacto series with *p* value 2.07E-05 and Fatty acid biosynthesis with *p* value 3.38E-22 are mapped from differentially co-expressed miRNAs extracted from GSE16759 AD dataset using WGCNA.

v. *Comparison:* In Table 7 of Supplementary file, we report the pathways mapped from differentially co-expressed miRNA modules extracted from the discussed module extraction techniques. Mutations in genes or gene targets, such as *α*-syn, PINKI, MAPT, HTRAC2, LRRK2, and DJ-1 induce degradation of dopaminergic neurons, leading to neurodegenerative diseases, such as AD and PD. From the early pioneering work on different enzymes regulating the mutations of *α*-syn, we know that defects in fatty acid biosynthesis enhances toxicity of *α*-syn [36], leading to AD.

3.2.2.2. PD datasets

i *Performance of THD-Module Extractor:* The common pathway mapped to the differentially co-expressed miRNA modules extracted from GSE72962 and GSE77668 PD datasets using THD-Module Extractor is Fatty acid biosynthesis with *p* values 8.53E-43 and 2.81E-22, respectively.

ii. *Performance of DiffCoEx:* The common pathway mapped to the differentially co-expressed miRNA modules extracted from GSE72962

Table 6
Pathway enrichment of differentially co-expressed miRNAs.

Dataset/Disease	Method	Pathways	*p* value
GSE72962- PD	THD-Module Extractor	FoxO signaling pathway	7.25E-07
		Fatty acid biosynthesis	8.53E-43
		Steroid biosynthesis	1.26E-11
	DiffCoEx	Fatty acid biosynthesis	3.38E-22
		Fatty acid metabolism	1.88E-16
	MODA	Steroid biosynthesis	1.26E-11
		ECM-receptor interaction	1.26E-11
	WGCNA	Adherens junction	4.5E-03
		Lysine biosynthesis	3.58E-03

and GSE77668 PD datasets using DiffCoEx is Fatty acid biosynthesis with *p* values 3.38E-22 and 3.01E-20, respectively.

iii. *Performance of MODA:* For datasets, GSE72962 and GSE77668 the best pathways mapped to the differentially co-expressed miRNA modules extracted using MODA are ECM-receptor interaction and Fatty acid biosynthesis with *p* values 1.26E-11 and 1.08E-12, respectively.

iv. *Performance of WGCNA:* For datasets, GSE72962 and GSE77668 the best pathways mapped to the differentially co-expressed miRNA modules extracted using WGCNA are Lysine Biosynthesis and Fatty acid biosynthesis with *p* values 3.58E-03 and 1.18E-18, respectively.

v. *Comparison:* In Table 6, we report the pathways mapped from differentially co-expressed miRNA modules extracted from GSE72962 PD dataset and observe that for THD-Module Extractor and DiffCoEx, the best differentially co-expressed miRNA module in terms of lower *p* value, the associated pathway is Fatty acids biosynthesis pathway with *p* values 8.53E-43 and 3.38E-22, respectively. For WGCNA, a pathway namely Lysine biosynthesis, has a *p* value of 3.58E-03, which suggests that the pathway is biologically significant. Disruptions in ubiquitin regulatory molecules is a hallmark of neurodegenerative diseases such as AD and PD. Lysine biosynthesis and modification of lysine linked ubiquitination is associated with pathogenesis of PD [37]. On the other hand, the pathway associated with the best differentially co-expressed miRNA module from MODA is ECM-receptor interaction. Along with other molecules, ECM (Extracellular matrix) molecules and their receptors are implicated during pathogenesis of neurodegenerative diseases, such as AD and PD [38]. Therefore, the differentially co-expressed miRNA modules extracted using the four techniques are enriched with PD-related pathways with higher biological significance. Table 8 of the Supplementary file show the pathway enrichment for the differentially co-expressed miRNA modules extracted from GSE77668 PD dataset.

From the above discussion, we infer that the miRNA modules extracted from AD and PD disease datasets are also associated with biologically significant pathways or biological processes.

3.3. Identification of hub genes across control and disease conditions

The gene or miRNA with highest degree of connecting genes or miRNAs in a module is called the hub gene or the hub miRNA for that module (Fig. 6). We extend differential co-expression analysis by identifying the hub gene(s) or miRNA(s) from each differentially co-expressed gene or miRNA modules extracted using the four methods. The hub gene(s) and the hub miRNA(s) has (have) highest degree(s) in gCEN and m_i CEN and are related with high biological significance. In Table 1 of Supplementary file, we list the hub genes identified from control and disease networks of differentially co-expressed gene modules of GSE68719-PD. Further, from downstream analysis of hub genes, it is observed that genes, such as RAD52, STPG1, SEMA3F, DPM1, CFH, BAD, CYP26B1, KLHL13, SCYL3, ANKIB1, LAP3, SARM1,

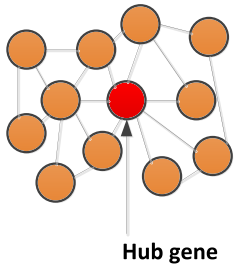

Fig. 6. A differentially co-expressed module with genes or miRNAs as the nodes. The red node with maximum number of directly connected neighboring nodes is the hub node for the module.

NDUFAB1,CDC27, LASP1, M6PR, WNT16, and KRIT1 lose their co-ordination with other genes across disease conditions. Further analysis of these hub genes has proved that these hub genes become inactive or undergo variations across disease conditions, thereby leading to neuronal cell death and neurodegenerative diseases, such as PD, AD, and multiple sclerosis (MS). For example, RAD52 is the principal candidate gene for DNA repair and protection of telomere structure. Inactivation or perturbation in RAD52 leads to cellular damage and neuronal death during aging [39]. Similarly, WNT signalling plays an important role in cellular functions, such as cell differentiation and in healthy functioning of the adult human brain. Dysregulation of WNT signalling or variations of genes regulating WNT signalling pathway, such as WNT16 suggests the pathogenesis of PD.

Table 2 of Supplementary file shows the changes in co-ordination among the hub miRNAs (the miRNAs with highest degrees) across control and disease m_i CENs of the differentially co-expressed miRNA modules of GSE72962-PD. Hub miRNAs such as hsa-miR-100-5p, hsa-miR-103a-3p, hsa-miR-106a-5p, hsa-miR-125b-5p, hsa-miR-1266-5p,hsa-miR-105-3p, hsa-miR-155-5p, hsa-miR-105-5p, hsa-let-7i-5p, hsa-miR-103a-2-5p, hsa-miR-103a-3p, hsa-miR-107, hsa-miR-100-3p, hsa-miR-34a-5p, hsa-miR-101-3p, hsa-miR-34c-3p, hsa-miR-34b, and hsa-miR-101-5p no longer have the highest degrees in the disease networks of the respective differentially co-expressed miRNAs. Interestingly, miRNAs such as hsa-miR-155-5p, hsa-miR-34b, and hsa-miR-34c-3p are found to be downregulated during progression of PD [2].

Therefore, from the above discussion, we conclude that the differentially co-expressed gene and miRNA modules are enriched with hub genes and hub miRNAs, which undergo variations in the network topology under disease conditions. Such dysregulation of hub genes and miRNAs is a hallmark of the underlying mechanism of PD.

3.4. Gene regulatory network analysis of hub genes

From differentially co-expressed genes, we find the hub genes which have the highest degrees in the corresponding disease networks. In Table 1 of Supplementary file, we list hub genes in the control network that lose their correlations with other hub genes in the disease network. In this section, we assess regulatory networks of the hub genes, mediated by the transciption factors. This regulatory network, called a gene regulatory network (GRN) provides insights on gene-gene interactions and helps identify and prioritize therapeutic targets in a disease. The regulatory genes govern the gene expression of mRNA and other proteins, which play vital roles in morphogenesis and pathogenesis of a disease. We use GENIE3 [40] to infer a regulatory network from expression data using an ensemble of Random forest of regression trees. Table 3 of Supplementary file presents the adjacency matrices of the GRNs of the hub genes, extracted from the GSE4757 AD dataset using THD-Module Extractor, DiffCoEx, MODA, and WGCNA. The hub genes such as DDR1, ABCA1, INPP5D, SLC24A4, ADRA1D, CDH2, CLEC7A, EGFR, ERAP1, GNAS, HLA-DRB1, PVR, ZNF423, BIN1, PTK2B, COPS5, IFNAR2, ABCA7, ABI3, ADAM10, ADAMTS4, ALPK2, ANKRD31, APH1B, APOE, BACE1, BACE2, BLOC1S3, CASS4, and CD2AP are found

Table 7
$Z_{Summary}$ scores based on permutation in AD and control conditions networks (medianRank>2) for differentially coexpressed modules from AD (GSE4757) dataset.

AD condition network			Control condition network		
Modules	medianRank	Zsummary	Modules	medianRank	Zsummary
black	12	1.7	black	10	4.3
yellow	11	10	yellow	6	12
brown	9	7.9	brown	4	16
red	7	6.3	red	7	10
magenta	6	16	magenta	5	8.4
blue	6	18	blue	1	57
turquoise	5	22	turquoise	7	11
pink	2	13	pink	9	6.6
green	2	28	green	2	17

to be the regulatory genes in the GRNs of the biomarkers, which have been validated and found to be associated with AD [41]. In Table 3 of Supplementary file, the weight of each regulatory gene signifies the link from the regulatory gene to a target gene.

3.5. Evaluation of module preservation statistics

Profiling of hub genes and hub miRNAs in control and disease conditions as discussed in Section 3.3 and 3.4, shows that differentially co-expressed genes and miRNAs do not preserve their network statistics (degree) across disease conditions. In other words, differentially co-expressed gene or miRNA modules are condition-specific and undergo regulatory changes and differences in gene connectivity between control and disease conditions. We use module preservation statistics to explore variations in topology of the gene co-expression network (gCEN) during progression of AD. In general, module preservation statistics [42] are used to find whether gene modules are preserved between reference and test networks. However, similar to Ref. [43], we use *modulePreservation()* to associate the perturbations and dysregulations of the disease network to the control network. Table 7 presents preservation statistics of a differentially co-expressed gene module, extracted from the AD (GSE4757) dataset using WGCNA. Lower preservation of differentially co-expressed genes shown in black, yellow, brown, and red modules depicts loss in co-expression structure between these gene pairs under AD conditions. Similarly in Fig. 7 (B), we identify that $Z_{Summary}$ for these four gene modules shown in black, yellow, brown, and red are between 10 and 2, which signifies that the modules are moderately preserved across the disease (AD) network.

From the experimental results given above, we list the main findings as follows:

- Differentially co-expressed gene modules with higher statistical significance (lower p and q values) may not always have better topological enrichment.
- Differentially co-expressed miRNA modules with higher statistical significance (lower p value) may not always have higher O/E ratio.
- Differentially co-expressed gene and miRNA modules, which are statistically significant are also mapped to pathways associated to AD or PD.
- Differentially co-expressed gene and miRNA modules are enriched with hub genes and hub miRNAs, respectively. The modules undergo variations in the network topology in disease conditions.
- A few of the hub genes identified in differentially co-expressed gene modules are seen to lose their correlations with other hub genes in the disease network.
- A few of the hub genes from the differentially co-expressed genes with significant weights in the GRN are found to be the regulatory genes of other AD related genes.
- Differentially co-expressed gene or miRNA modules are condition-specific and undergo regulatory changes and differences in gene connectivity between control and disease conditions.

4. Discussion

In view of the performance of each method, we discuss the strengths and weaknesses of the methods as follows.

- Among the four discussed techniques, WGCNA is the most widely used for CEN module extraction and is well supported and well documented for analysis of CEN modules for module preservation, and hub gene identification. Except WGCNA, none of the CEN modules detected are used with RNA-seq data. Therefore, in view of understanding the behavior of differentially co-expressed modules, we apply the four techniques to both gene and miRNA expression profiles of microarray and RNA-seq datasets. The differentially co-expressed genes and miRNAs modules are enriched with GO terms, biological pathways, and other miRNA-gene targets involved in the progression of a disease.
- Unlike WGCNA, THD-Module Extractor does not impose limit on the number of samples in a dataset. Therefore, THD-Module Extractor also performs well in extracting differentially co-expressed miRNA modules, with more significant p values and better miRNA-target enrichment for both GSE16759 and GSE77668 datasets, which have fewer sample sizes. Another strong argument which supports the superior performance of THD-Module Extractor in terms of GO enrichment and topological significance is that unlike other CEN module extraction techniques, THD-Module Extractor considers high semantic similarity of border genes in detecting co-expressed modules.
- On the other hand, unlike other CEN module extraction techniques, THD-Module Extractor is implemented in both MATLAB and *R*.
- Both MODA and DiffCoEx are also useful in detecting differentially co-expressed genes and miRNAs from CEN in addition to detecting co-expressed modules with high correlations among the entities.

We also assess the differentially co-expressed modules in terms of pathway enrichment, miRNA target enrichment, and hub gene identification for exploring the potential roles in disease pathogenesis. GO terms, such as cellular nitrogen compound metabolic process, ion binding, and gene expression with lower p values are involved in slow progression of AD. Similarly, evidence from the past has suggested that pathways, namely Fatty acid biosynthesis, Lysine biosynthesis, and ECM-receptor interaction pathways are also related to PD. miRNA targets such as A2BAR, HTR1B, SLC5A11, WNT16, PINK1, RAD52, and PLA2G3, showing higher O/E ratios, are found to have been involved in pathogenesis of neurodegenerative diseases, such as AD and PD.

In a gCEN or m_i CEN network, the core gene(s) or core miRNA(s) with highest degree(s) is (are) the principal and potential gene(s), which may undergo variations across disease conditions. Therefore, analysis of the hub gene(s) or miRNA(s) from each differentially co-expressed gene or miRNA module, predicts the potential perturbations of the regulatory mechanisms of a disease. We found that hub gene(s) such as RAD52 and WNT16, which lose their co-regulation with other genes across disease conditions are the key regulators of cellular functions, such as cell repair, and cell differentiation. Loss of linkages of these hub genes in a disease network leads to neuronal cell death and progression of PD. Similarly, hub miRNAs such as hsa-miR-155-5p, hsa-miR-34b, and hsa-miR-34c-3p are downregulated during disease conditions, which implicate them in their roles in the progression of PD.

We further examine the global differences among the differentially co-expressed genes in disease mechanisms underlying PD and AD using *modulePreservation()*, given by Ref. [42]. We found that the differentially co-expressed modules show a poor evidence of preservation of $Z_{Summary}$, keeping the disease network as reference network and the control network as the test network. Therefore, from the above

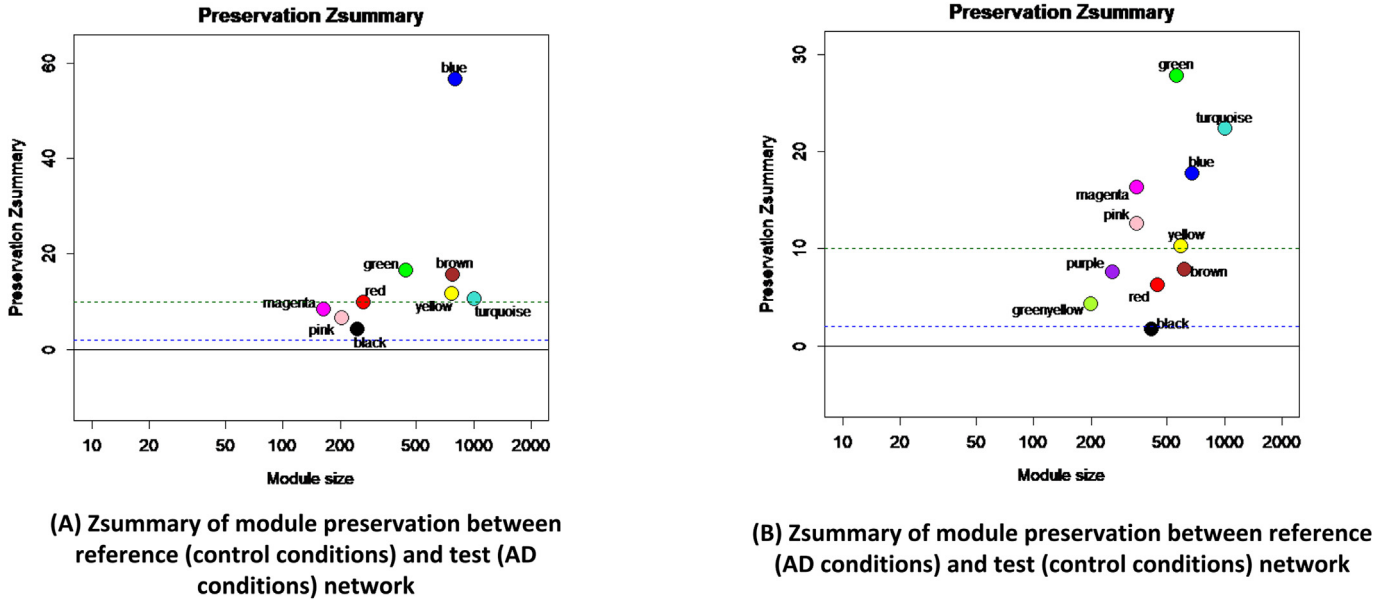

Fig. 7. Scatter plot with $Z_{Summary}$ of module preservation between reference and test networks of differentially co-expressed gene modules extracted from AD (GSE4757) dataset. (A) shows the preservation $Z_{Summary}$ of the differentially co-expressed gene modules taking control conditions as reference and AD conditions as test networks. (B) shows the preservation $Z_{Summary}$ of the differentially co-expressed gene modules taking AD conditions as reference and control conditions as test networks.

discussion, we can infer that the differentially co-expressed genes or miRNAs extracted using the four methods are condition-specific and points to dysfuctional interactions and pathways.

5. Conclusion

Alzheimer's Disease (AD) and Parkinson's Disease (PD), two deadly neurodegenerative diseases are associated with decline of memory and other voluntary functions, and are commonly seen in elderly persons. Analysis of AD and PD datasets using high throughput technologies such as microarray and next-generation sequencing (NGS) data can provide better insights to identification of disease related biomarkers. In step with the recent research leaps in high throughput technologies, there is also advancement in computational methods and approaches, including module extraction techniques to mine the disease related biomarkers. To the best of our knowledge, this is the first empirical study reviewing the leading module extraction techniques, namely THD-Module Extractor, DiffCoEx, MODA, and WGCNA to identify differentially coexpressed gene and miRNA modules and analyze the differences in gene connectivity patterns between conditions using both microarray and RNA-seq data for AD and PD. Though there are limitations in preprocessing and analysis of RNA-seq data, the throughput of the analysis is quite informative. This work is a guide cum empirical study of the CEN techniques widely used in the research domain to identify the differentially co-expressed genes and miRNA modules.

Based on the detailed analysis, we conclude that all four methods produce biologically enriched and statistically significant differentially co-expressed modules of genes and miRNAs. However, the differentially co-expressed gene and miRNA modules extracted using THD-Module Extractor outperform in comparison to modules found by the other three discussed methods in terms of GO enrichment, topological significance, pathway analysis, hub gene identification, gene regulatory network analysis, and disease biomarker identification. The working principle of THD-Module Extractor is that it includes both gene expression similarity and semantic similarity to identify co-expressed modules and this results in the mining of differentially co-expressed gene or miRNA modules, including functionally significant genes or miRNAs, associated with AD or PD related pathways. However, there is a want of a web-based application for the ease of use and reproducibility of THD-Module Extractor.

We believe that disease biomarker identification using up-to-date computational techniques assists in making sense of ever-growing amounts of genetic data associated with various debilitating diseases that are becoming available, so that the findings can be used to bring benefits to human health. Therefore, by discussing and comparing the capabilities of several module extraction techniques, we hope to assist researchers in discovering and prioritizing potential disease biomarkers for further investigation.

Author contributions & funding

All authors contributed in the design of the framework. TK implemented the method using R, and wrote the manuscript. DKB, PB, and JK reviewed the manuscript.

There is no source of funding.

Conflicts of interest

None Declared.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compbiomed.2019.103380.

References

[1] Sanghamitra Bandyopadhyay, Malay Bhattacharyya, Analyzing miRNA co-expression networks to explore tf-miRNA regulation, BMC Bioinf. 10 (1) (2009) 163.
[2] Edoardo Missiaglia, Chris J. Shepherd, Ewa Aladowicz, David Olmos, Joanna Selfe, Gaëlle Pierron, Olivier Delattre, Zoe Walters, Janet Shipley, MicroRNA and gene co-expression networks characterize biological and clinical behavior of rhabdomyosarcomas, Cancer Lett. 385 (2017) 251–260.
[3] Tulika Kakati, Hirak Kashyap, Dhruba K. Bhattacharyya, THD-module extractor: an application for CEN module extraction and interesting gene identification for alzheimers disease, Sci. Rep. 6 (2016) 38046.
[4] Bruno M. Tesson, Rainer Breitling, Ritsert C. Jansen, Diffcoex: a simple and sensitive method to find differentially coexpressed gene modules, BMC Bioinf. 11 (1) (2010) 497.

[5] Dong Li, James B. Brown, Luisa Orsini, Zhisong Pan, Guyu Hu, He Shan, MODA: module differential analysis for weighted gene co-expression network, 1605.04739, (2016).
[6] Peter Langfelder, Steve Horvath, WGCNA: an R package for weighted correlation network analysis, BMC Bioinf. 9 (1) (2008) 559.
[7] S. van Dam, U. Võsa, A. van der Graaf, L. Franke, J.P. de Magalhães, Gene co-expression analysis for functional classification and gene–disease predictions, Briefings Bioinf. 19 (2017) 575–592.
[8] Thomas WH. Lui, Nancy BY. Tsui, Lawrence WC. Chan, Cesar SC. Wong, Parco MF. Siu, Benjamin YM. Yung, DECODE: an integrated differential co-expression and differential expression analysis of gene expression data, BMC Bioinf. 16 (1) (2015) 182.
[9] Ana Conesa, Pedro Madrigal, Sonia Tarazona, David Gomez-Cabrero, Alejandra Cervera, Andrew McPherson, Michał Wojciech Szcześniak, Daniel J. Gaffney, Laura L. Elo, Xuegong Zhang, et al., A survey of best practices for RNA-seq data analysis, Genome Biol. 17 (1) (2016) 13.
[10] Lisette JA. Kogelman, Susanna Cirera, Daria V. Zhernakova, Merete Fredholm, Lude Franke, Haja N. Kadarmideen, Identification of co-expression gene networks, regulatory genes and pathways for obesity based on adipose tissue RNA sequencing in a porcine model, BMC Med. Genomics 7 (1) (2014) 57.
[11] Jae-Hyung Lee, Chen Gao, Guangdun Peng, Christopher Greer, Shuxun Ren, Yibin Wang, Xinshu Xiao, Analysis of transcriptome complexity through rna sequencing in normal and failing murine heartsnovelty and significance, Circ. Res. 109 (12) (2011) 1332–1341.
[12] Lihua Wang, Zhi-Ping Liu, Detecting diagnostic biomarkers of alzheimer's disease by integrating gene expression data in six brain regions, Front. Genet. 10 (2019) 157–157.
[13] Hussain Ahmed Chowdhury, Dhruba Kumar Bhattacharyya, Jugal K. Kalita, (Differential) Co-expression analysis of gene expression: a survey of best practices, IEEE ACM Trans. Comput. Biol. Bioinform (2019).
[14] Bao-Hong Liu, Differential coexpression network analysis for gene expression data, Computational Systems Biology, Springer, 2018, pp. 155–165.
[15] Hasin Afzal Ahmed, Priyakshi Mahanta, Dhruba Kumar Bhattacharyya, Jugal Kumar Kalita, Shifting-and-scaling correlation based biclustering algorithm, IEEE ACM Trans. Comput. Biol. Bioinform 11 (6) (2014) 1239–1252.
[16] Erzsébet Ravasz, Anna Lisa Somera, Dale A. Mongru, Zoltán N. Oltvai, A.-L. Barabási, Hierarchical organization of modularity in metabolic networks, Science 297 (5586) (2002) 1551–1555.
[17] Paulami Chatterjee, Debjani Roy, Malay Bhattacharyya, Sanghamitra Bandyopadhyay, Biological networks in Parkinsons disease: an insight into the epigenetic mechanisms associated with this disease, BMC Genomics 18 (1) (2017) 721.
[18] Elena Atrián-Blasco, Amandine Conte-Daban, Christelle Hureau, Mutual interference of Cu and Zn ions in Alzheimer's disease: perspectives at the molecular level, Dalton Trans. 46 (38) (2017) 12750–12759.
[19] Giancarlo V. De Ferrari, Nibaldo C. Inestrosa, Wnt signaling function in alzheimer's disease, Brain Res. Rev. 33 (1) (2000) 1–12.
[20] Bart De Strooper, Robert Vassar, Todd Golde, The secretases: enzymes with therapeutic potential in alzheimer disease, Nat. Rev. Neurol. 6 (2) (2010) 99.
[21] Chun-Qiu Dai, Ting-Ting Luo, Shi-Cheng Luo, Jia-Qi Wang, Sheng-Ming Wang, Yun-Hu Bai, Yan-Ling Yang, Ya-Yun Wang, p53 and mitochondrial dysfunction: novel insight of neurodegenerative diseases, J. Bioenerg. Biomembr. 48 (4) (2016) 337–347.
[22] Ha-Na Woo, Jong-Sung Park, A-Ryeong Gwon, Thiruma V. Arumugam, Dong-Gyu Jo, Alzheimers disease and notch signaling, Biochem. Biophys. Res. Commun. 390 (4) (2009) 1093–1097.
[23] Norikazu Kiguchi, Yuka Kobayashi, Shiroh Kishioka, Chemokines and cytokines in neuroinflammation leading to neuropathic pain, Curr. Opin. Pharmacol. 12 (1) (2012) 55–61.
[24] Tzu-Pin Lu, Chien-Yueh Lee, Mong-Hsun Tsai, Yu-Chiao Chiu, Chuhsing Kate Hsiao, Liang-Chuan Lai, Eric Y. Chuang, miRSystem: an integrated system for characterizing enriched functions and pathways of microRNA targets, PLoS One 7 (8) (2012) e42390.
[25] Wei Kong, Xiaoyang Mou, Jin Deng, Benteng Di, Ruxing Zhong, Shuaiqun Wang, Yang Yang, Weiming Zeng, Differences of immune disorders between alzheimers disease and breast cancer based on transcriptional regulation, PLoS One 12 (7) (2017) e0180337.
[26] Kenji Takami, Akinori Matsuo, Kazuhiro Terai, Douglas G. Walker, Edith G. McGeer, Patrick L. McGeer, Fibroblast growth factor receptor-1 expression in the cortex and hippocampus in Alzheimer's disease, Brain Res. 802 (1–2) (1998) 89–97.
[27] Tara Leigh Spires-Jones, Leora M. Fox, Anete Rozkalne, Rose Pitstick, George A. Carlson, Aleksey G. Kazantsev, Inhibition of sirtuin 2 with sulfobenzoic acid derivative AK1 is non-toxic and potentially neuroprotective in a mouse model of frontotemporal dementia, Front. Pharmacol. 3 (2012) 42.
[28] Rene O. Sanchez-Mejia, Lennart Mucke, Phospholipase a 2 and arachidonic acid in Alzheimer's disease, Biochim. Biophys. Acta Mol. Cell Biol. Lipids 1801 (8) (2010) 784–790.
[29] Jungsu Kim, Jacob M. Basak, David M. Holtzman, The role of apolipoprotein e in Alzheimer's disease, Neuron 63 (3) (2009) 287–303.
[30] Hyejin Park, Young Doo Kim, Tae-In Kam, Hyunwoo Choi, Changsoo Kim, Yong-Keun Jung, Role of AK1 in tau phosphorylation and toxicity, Alzheimer's Dementia: J. Alzheimer's Assoc. 6 (4) (2010) S277.
[31] Dubravka Švob Štrac, Nela Pivac, Dorotea Mück-Šeler, The serotonergic system and cognitive function, Transl. Neurosci. 7 (1) (2016) 35–49.
[32] Syed Faraz Kazim, Khalid Iqbal, Neurotrophic factor small-molecule mimetics mediated neuroregeneration and synaptic repair: emerging therapeutic modality for Alzheimers disease, Mol. Neurodegener. 11 (1) (2016) 50.
[33] Stefania Gessi, Stefania Merighi, Katia Varani, Pier Andrea Borea, Adenosine receptors in health and disease, Advances in Pharmacology, vol. 61, Elsevier, 2011, pp. 41–75.
[34] Rangarajan Parakalan, Boran Jiang, Baby Nimmi, Manivannan Janani, Manikandan Jayapal, Jia Lu, Samuel SW. Tay, Eng-Ang Ling, S Thameem Dheen, Transcriptome analysis of amoeboid and ramified microglia isolated from the corpus callosum of rat brain, BMC Neurosci. 13 (1) (2012) 64.
[35] Abel Ickowicz, Yu Feng, Karen Wigg, Jennifer Quist, Tejasawe Pathare, Wendy Roberts, Molly Malone, Russell Schachar, Rosemary Tannock, James L. Kennedy, et al., The serotonin receptor HTR1B: gene polymorphisms in attention deficit hyperactivity disorder, Am. J. Med. Genet. Part B: Neuropsychiatric Genetics 144 (1) (2007) 121–125.
[36] Yong Joo Lee, Shaoxiao Wang, Sunny R. Slone, Talene A. Yacoubian, Stephan N. Witt, Defects in very long chain fatty acid synthesis enhance alpha-synuclein toxicity in a yeast model of Parkinson's disease, PLoS One 6 (1) (2011) e15946.
[37] Jeanne MM. Tan, Esther SP. Wong, Donald S. Kirkpatrick, Olga Pletnikova, Han Seok Ko, Shiam-Peng Tay, Michelle WL. Ho, Juan Troncoso, Steven P. Gygi, Michael K. Lee, et al., Lysine 63-linked ubiquitination promotes the formation and autophagic clearance of protein inclusions associated with neurodegenerative diseases, Hum. Mol. Genet. 17 (3) (2007) 431–439.
[38] Vladimir Berezin, Peter S. Walmod, Mikhail Filippov, Alexander Dityatev, Targeting of ECM molecules and their metabolizing enzymes and receptors for the treatment of CNS diseases, Progress in Brain Research, vol. 214, Elsevier, 2014, pp. 353–388.
[39] Silvia Mercado-Sáenz, Beatriz López-Díaz, Francisco Sendra-Portero, Manuel Martínez-Morillo, Miguel J. Ruiz-Gómez, Inactivation of RAD52 and HDF1 DNA repair genes leads to premature chronological aging and cellular instability, J. Biosci. 42 (2) (2017) 219–230.
[40] Alexandre Irrthum, Louis Wehenkel, Pierre Geurts, et al., Inferring regulatory networks from expression data using tree-based methods, PLoS One 5 (9) (2010) e12776.
[41] Yan-Shi Hu, Juncai Xin, Ying Hu, Lei Zhang, Ju Wang, Analyzing the genes related to Alzheimers disease via a network and pathway-based approach, Alzheimer's Res. Ther. 9 (1) (2017) 29.
[42] Peter Langfelder, Steve Horvath, Eigengene networks for studying the relationships between co-expression modules, BMC Syst. Biol. 1 (1) (2007) 54.
[43] Mohit Jha, Anvita Gupta, Sudha Singh, Khushhali Menaria Pandey, Differential co-expression network analysis for identification of functional modules in hiv infection and tuberculosis-hiv co-infection, J. Biol. Res. Bollettino della Società Italiana di Biologia Sperimentale 89 (2) (2016).

Tulika Kakati was born in Assam, India in 1991. She received her BE in Computer Science & Engineering from Jorhat Engineering College in 2013 and M.Tech in Information Technology in 2015 from Tezpur University. Currently, she is pursuing Ph.D. in Tezpur University. Her current research interests include Machine Learning and Computational Biology.

Dhruba Kumar Bhattacharyya received his Ph.D. in Computer Science from Tezpur University in 1999. He is a Professor in the Computer Science & Engineering at Tezpur University. His research areas include Data Mining, Network Security and Content-based Image Retrieval. Prof. Bhattacharyya has published 250 + research papers in leading international journals and conference proceedings. In addition, Dr Bhattacharyya has written and edited many books. He is Program Committee and Advisory Body member of several international conferences/workshops.

Pankaj Barah is currently an Assistant Professor and Ramalingaswami Re-Entry Fellow in the Department of Molecular Biology and Biotechnology (MBBT) in Tezpur University. He has received his M.Sc. degree in Bioinformatics (2006) from University of Madras in India and PhD in Computational Systems Biology (2013) from the Norwegian University of Science and Technology, Trondheim, Norway. He has worked as Bioinformatics scientist in the division of Theoretical Bioinformatics at German Cancer Research Center (DKFZ) in Heidelberg, Germany during 2015–2017. His research areas include-computational systems biology, bioinformatics, evolutionary systems biology, Next Generation Sequencing (NGS), Big data analytics and biological networks.

Jugal Kumar Kalita is a professor of Computer Science at the University of Colorado at Colorado Springs. He received his Ph.D. from the University of Pennsylvania in 1990. His research interests are in Natural Language Processing, Artificial Intelligence and Bioinformatics. He has published more than 200 papers in international journals and referred conferences.