Biochemical and Biophysical Research Communications 484 (2017) 557–564

Contents lists available at ScienceDirect

Biochemical and Biophysical Research Communications

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

Comparative analysis of RNA-Seq data from brain and blood samples of Parkinson's disease

Paulami Chatterjee, Debjani Roy*

Department of Biophysics, Bose Institute, Acharya J.C. Bose Centenary Building, P-1/12 C.I.T Scheme VII M, Kolkata, 700054 West Bengal, India

ARTICLE INFO

Article history:
Received 18 January 2017
Accepted 23 January 2017
Available online 26 January 2017

Keywords:
Parkinson's disease
RNA-Seq data
miR Co-expression network
Pathway analysis
SM2miR drugs

ABSTRACT

Parkinson's disease (PD) is the second most common neurodegenerative disorders throughout the world. In order to search for PD biomarkers, we performed a system-level study of RNA-Seq data from PD brain and blood samples. Differentially expressed miRs of RNA-Seq data were subjected to generate the Co-expression networks. Three highly co-expressed clusters were identified based on their correlation co-efficient values and fold change ratio. SM2miR drugs of the miRs contained in the three highly co-expressed clusters were identified, and drugs common among these clusters were selected. Co-expressed miRs not previously known to be associated with PD were identified from both the samples. Functional enrichment analyses of these miR targets were done, and the pathways common and unique to both the samples were identified. Thus, our study presents a comparative analysis of miRs, their associated pathways, and drugs from brain and blood samples of PD that may help in system level understanding of this disease. miRs identified from our study may serve as biomarkers for PD.

1. Introduction

Parkinson's disease (PD) is the second most common neurodegenerative disorders throughout the world. It is one of the chronic and progressive movement disorders in which symptoms get worse over time. Deregulation of microRNAs (miRs) has been implicated as one of the major causative factors in neurodegenerative diseases [1], including PD [2–4]. The previous study has pointed out that few miRs can be predominantly expressed in certain tissues, whereas the majority of miRs (82.9%) are neither specific for single tissues nor housekeeping miRs [5]. Therefore, knowledge of the expression pattern and distribution of different miRs in tissues is essential for the understanding of the physiological and pathological mechanisms associated with a disease [5]. Analyses of genome-wide small RNA transcriptome provide an accurate and comprehensive knowledge of miR expression in the disease under different conditions [6].

RNA sequencing (RNA-Seq) which uses Next Generation Sequencing technology to quantify and discover RNAs in a cellular transcriptome [7,8], has been widely used for investigating the miR-mediated pathogenesis of complex diseases. The high-resolution, genome-wide miR profile of PD obtained from RNA-seq offer insight into the molecular and pathological mechanisms that occur in the disease [9]. RNA-Seq mediated comparative blood transcriptome analysis of idiopathic and LRRK2 mutated PD has resulted in the identification of the significant difference in peripheral blood transcriptome [10]. Long non-coding RNAs and alternative splicing modulations have been identified from PD blood cells by RNA-Seq method [11].

However, the comparative analysis of miR profiles of RNA-Seq data from brain and blood samples has not been performed yet for PD. In this computational study, starting from PD brain and blood RNA-Seq data, we have identified disease-specific highly co-expressed miRs. Of these co-expressed miRs, we screened the miRs that are not previously known to be associated with PD. In this work, we studied highly co-expressed miRs, their associated pathways, and drugs from brain and blood samples that may help in system level understanding of this disease. The miRs identified from our study may serve as biomarkers for PD.

2. Methodology

Fig. 1 depicts the workflow of our methodology.

Abbreviations: PD, Parkinson's Disease; miR, MicroRNA; RNA-Seq, RNA sequencing; DE, Differentially Expressed; Hclust, Hierarchical Clustering; ATC, Anatomic Therapeutic Classification.

* Corresponding author. Department of Biophysics, Bose Institute, Acharya J.C. Bose Centenary Building, P-1/12 C.I.T Scheme VII M, Kolkata, 700054, India.
E-mail addresses: chatterjee_paulami@yahoo.co.in (P. Chatterjee), roydebjani13@gmail.com (D. Roy).

http://dx.doi.org/10.1016/j.bbrc.2017.01.121
0006-291X/

2.1. Data collection

RNA-Seq data of brain and blood samples of PD were retrieved from Gene Expression Omnibus (GEO) [Accession IDs- GSE72962 and GSE40915]. GSE72962 (Data 1) contains 29 PD and 33 neuropathologically-normal control from post-mortem prefrontal cortex Brodmann Area 9 (BA9) [9]. GSE40915 (Data 2) contains 7 PD patients before and after Deep Brain Simulation (DBS) treatment (which improves some of the PD motor symptoms) and 6 age-matched healthy control samples from blood leukocytes [6].

2.2. Differentially expressed miR selection

In order to identify the differentially expressed (DE) miRs over disease and normal conditions, paired two sample *t*-test was applied on each of the expression datasets. The *t*-test measures the statistical significance of the dataset in terms of test statistic *t*, which is given by:

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{m}}}$$

Where $\bar{x}$ and $\bar{y}$ are the sample means, s_x and s_y are the sample standard deviations, n and m are the sample sizes for two samples, x and y. Under the null hypothesis, this test returns the probability (p-value) of observing a value as extreme or more extreme of the

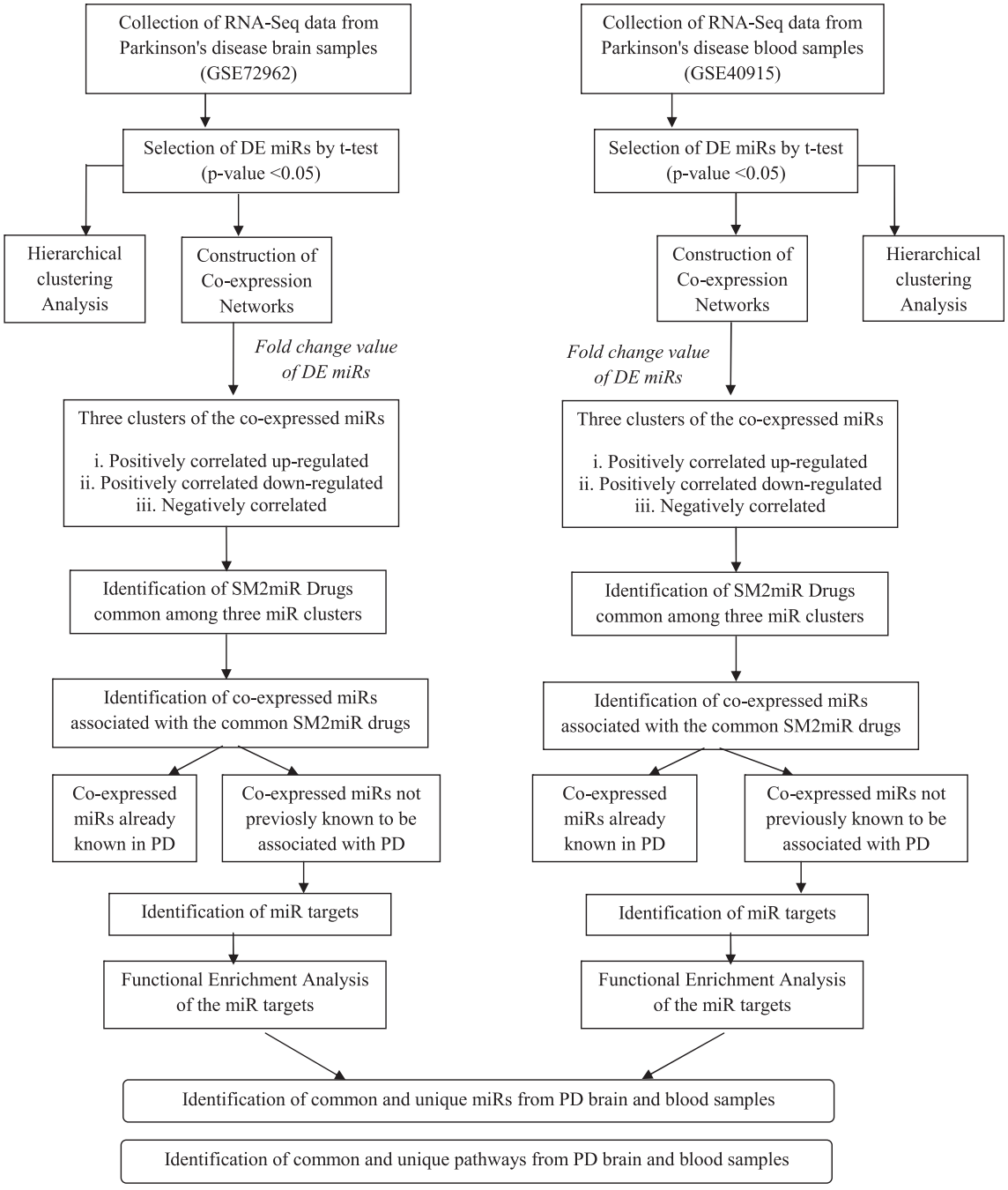

Fig. 1. Methodology for the analysis of RNA-Seq data from brain and blood samples of PD.