Gene 584 (2016) 167–172

Contents lists available at ScienceDirect

Gene

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

Research Paper

Dysregulation of miRNA isoform level at 5′ end in Alzheimer's disease

Shengqin Wang [a,b,1], Yuming Xu [b,1], Musheng Li [b], Jing Tu [b], Zuhong Lu [b,c,*]

[a] College of Life and Environmental Science, Wenzhou University, Wenzhou 325035, China
[b] State Key Lab of Bioelectronics, School of Biological Science and Medical Engineering, Southeast University, Nanjing 210096, China
[c] Department of Biomedical Engineering, College of Engineering, Peking University, Beijing 100781, China

ARTICLE INFO

Article history:
Received 26 July 2015
Received in revised form 9 January 2016
Accepted 15 February 2016
Available online 17 February 2016

Keywords:
Dysregulation
miRNA
Isoform level
IsomiR
Alzheimer's disease

ABSTRACT

Alzheimer's disease (AD) is the most common form of dementia, whose mechanism is still not yet fully understood. A miRNA-based signature method, commonly according to the changes of expression levels, is widely used for AD analysis in previous studies. Recently, miRNA isoforms called as isomiR variants, which is considered to play important biological roles, have been demonstrated as the applications of high throughput sequencing platforms. Here, we presented an entropy-based model to detect the miRNA isoform level at the 5′ end, and found many miRNAs with significant changes of isoform levels between the early stage and the late stage of AD by the application of this model to the public data. The statistical significance of the overlap between isoform-level changed miRNAs and AD related miRNAs extracted from HMDD2 supports that these miRNA isoforms are not degradation products. Based on the most common isomiR seed analysis of isoform-level changed AD related miRNAs, the predicted targets are also found to be enriched for genes involved in transcriptional regulation and the nervous system. After comparing with the expression level based method, we detected that changes of 5′ isoform levels are more stable than those of expression levels for AD related miRNA detecting.

1. Introduction

Alzheimer's disease (AD) is a complicated progressive neurodegenerative disease leading to dementia. Based on the stochastic, multistate model, one in 85 persons worldwide will be affected by 2050, and delaying the onset and the progression of AD can dramatically decrease its global burden (Brookmeyer et al., 2007).

The evidence for the genetic component in AD shows the complexity pathophysiology of this disease (Maes et al., 2009; Ertekin-Taner, 2007; Hardy, 2006), which may link metabolic and pathologic pathways, and the miRNAs involved in the etiopathogenetic pathways may be also anticipated. Actually, a number of deregulated miRNA have been identified to correlate with AD (Delay et al., 2012; Nelson and Wang, 2010). MiRNAs/microRNAs are small non-coding RNAs that function as guide molecules in RNA silencing in the post-transcriptional regulation network by targeting most mRNA genes (Bartel, 2004; Huntzinger and Izaurralde, 2011). Under the guidance of the RISC (RNA-induced silencing complex), miRNAs can anneal to mRNA with partial complementary target sequence (Kim et al., 2009; Grimson et al., 2007). Recently, based on the data from microarray and next generation sequencing platform, most of studies have focused on the changes of miRNA expression levels in AD (Geekiyanage et al., 2012; Leidinger et al., 2013; Kumar et al., 2013).

In animals, a miRNA is ~22 nucleotides in length and typically processed from a primary miRNA transcript (pri-miRNA) by a series of enzymes including Drosha in the nucleus and Dicer in the cytoplasm, releasing a short double-stranded RNA (dsRNA) duplex consisting of one or two functional mature miRNA (Gu et al., 2012; Ha and Kim, 2014). The variation in Drosha or Dicer processing and non-templated nucleotide variation can generate miRNA isoforms (isomiRs), different from the canonical mature miRNA by a few nucleotides at the 5′ or 3′ end and termed 5′ or 3′ isomiR (Neilsen et al., 2012). The existence of isomiR variants has been demonstrated by various next generation high throughput small RNA sequencing projects (Guo and Lu, 2010; Cloonan et al., 2011; Körbes et al., 2012; Pritchard et al., 2012). Frequent variations seen repeatedly at the same site are unlikely attributed to degradation or sequencing error. A number of groups have reported that miRNA variations involved in regulating distinctive target genes can play an important biological role in controlling miRNA-mediated gene expression (Vickers et al., 2013; Ameres and Zamore, 2013; Fernandez-Valverde et al., 2010; Guo et al., 2013; Hinton et al., 2014; Telonis et al., 2015).The variations in the 3′ end can alter the stability of miRNAs and the efficiency of target repression (Wyman et al., 2011; Katoh et al., 2009; Jones et al., 2009). The variants in the 5′ end leads to a shift of the canonical seed region (typically 2–7 nt of the miRNA) that is intended to be very critical to determine miRNA target specificity (Wang, 2014; Wang et al., 2014; Friedman et al., 2009), thus reshuffling

Abbreviations: AD, Alzheimer's disease; RISC, RNA-induced silencing complex; pri-miRNA, primary miRNA transcript; isomiRs, miRNA isoform products; FDR, false discovery rate.

* Corresponding author at: State Key Lab of Bioelectronics, School of Biological Science and Medical Engineering, Southeast University, Nanjing 210096, China.
E-mail address: zhlu@seu.edu.cn (Z. Lu).

[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.gene.2016.02.020
0378-1119/

the target repertoire (Tan et al., 2014; Bizuayehu et al., 2012; Ebhardt et al., 2010; Humphreys et al., 2012). Compared with the 3′ end, the 5′ end of miRNA should have higher biological function pressure.

Currently, two tools have been developed to detect the isomiRs: the isomiRex is designed to detect the miRNA isoform with higher expression than the canonical mature miRNA (Sablok et al., 2013), and isomiRID simply lists all candidate miRNA isoforms (de Oliveira et al., 2013). Since tools for modeling the isoform levels of specific isomiRs are still limited, the biological relevance of these molecules is largely unknown though a set of isomiRs has been experimentally detected (Neilsen et al., 2012). For example, a recent study could not determine the relationship between AD and the changes of miRNA variants by only analyzing the most common isomiR (Hébert et al., 2013). Here, we presented a novel entropy-based model *MIH5* (please find the detailed descriptions in the Methods section) to detect the dysregulation of miRNA isoform level at the 5′ end. It is found that many miRNAs had significant changes of isoform levels between the early stage and the late stage of AD, which means that the dysregulation of miRNA isoform level at the 5′ end should play an important role in AD.

2. Methods

2.1. Data sources and miRNA analysis

MiRNA next generation sequencing data were retrieved from the GEO database (http://www.ncbi.nlm.nih.gov/geo, accession number GSE48552) (Lau et al., 2013), which contained unpaired samples of 6 early stage AD and 6 later stage AD. Researchers in Belgium carried out a study of original experiments. In order to detect mature miRNAs from the sequencing data, we used the tool cutadapt to remove the adaptor sequence (Martin, 2011). Based on the pre-miRNA and canonical mature miRNA information downloaded from the mirBase database (Kozomara and Griffiths-Jones, 2014), we employed the isomiRID tool to detect the mature miRNA with variation. Briefly, the isomiRID used the bowtie to map reads to the pre-miRNAs to find out the perfect matched mature miRNAs, then filtered non-miRNA reads after mapping reads to the human genome and transcriptome. The remained reads were mapped to the pre-miRNA with 1 mismatch to find the non-template variation related miRNAs. Here, we set the parameter as: M3: no no and M5: no no, which means we do not detect the non-templated adding, in order to decrease the false positive result.

2.2. Entropy calculation

As described in our previous study, we defined the isomiRs as sequences matching the known pre-miRNA in the canonical mature miRNA region ±4 nt (Wang et al., 2015). To characterize the 5′ isomiR variants of a given miRNA in the case of aligned miRNA sequences, we defined *MIH*, MiRNA Isoform entropy (abbreviated H), as the average of Shannon entropies of the observed symbol distribution for each site of the canonical mature miRNA region ±4 nt, and *MIH5* as related score of the 5′ end region ±4 nt:

$$MIH5 = -\frac{1}{L}\sum_{i}^{L}\sum_{j}^{N} p_j^i \log_2 p_j^i. \quad (1)$$

Here, p_j^i is calculated as the frequency of the character j at the particular sequence position i divided by the number of sequences in the alignment. N is the number of distinct symbols for the given sequence type (four for RNA: A, U, G, and C; note that in order to maximize the ability to get the difference of entropy values, the gap character "-" is not included). Here, L is equal to 9, including the end site and flanking region (±4 nt). For example, as shown in Fig. 1, the isomiRs of hsa-miR-153-5p was defined as the non-canonical mature miRNA sequences located in the black dotted frame, and the corresponding *MIH5* could be calculated from the gray solid rectangle region. Here, the miRNAs for each sample with the count equal or greater than 50 were retained for further analysis.

2.3. Statistic test and validation data

In order to find out whether the dysregulation of 5′ isomiR expression is correlated with AD, we calculated the *MIH5* for miRNAs (count ≥50) in each sample and applied the two-sided Wilcoxon rank sum (non-parametric) test to infer the statistical significant difference between 6 early stage AD samples and 6 later stage AD samples. We calculated the P-values followed by the Bonferroni post hoc correction for each available miRNA gene.

In order to check whether the dysregulation of miRNA isoform level at the 5′ end is associated with AD, we compared our result with 30 known AD related miRNAs extracted from HMDD2.

2.4. IsomiR seeds analysis

For AD related miRNA with dysregulation of isoform level, we predicted the candidate function for the most common isomiRs. The entropy difference between the early stage and the later stage of AD can present the change of miRNA isoform level, especially the proportion of the most common isomiRs. Therefore, we extracted the most common isomiR seeds of AD related miRNAs found in this study with a significant difference of the *MIH5* to detect the candidate function. We predicted the candidate target gene by the TargetScan Custom (http://www.targetscan.org/vert_50/seedmatch.html), which can detect the biological targets of miRNAs based on the seed region. Then, the predicted target genes were directed to the DAVID functional annotation tools to implement the function enrichment analysis. Here, we choose the 3 Gene Ontology items (GOTERM_BP_FAT, GOTERM_CC_FAT, and GOTERM_MF_FAT) for functional annotation, and the Enrichment Thresholds or EASE was set as 0.01.

2.5. Differential expression of miRNAs

In order to compare with the results from expression level based methods, we employed the tools to find out the miRNAs with significant changes of expression levels. The R package DESeq (Anders and Huber, 2010) and edgeR (Robinson et al., 2010) are widely used to analyze the digital-based differential expression of RNA sequencing data, both of which detected the significantly changed miRNAs based on the negative binomial distribution model (Anders et al., 2013). Here, the number of specific mature miRNA was calculated by the reads that fall within the same position on the precursor, plus 4 nt upstream and 4 nt downstream. The edgeR package was used to normalize the miRNA data and calculate the fold change. Both the edgeR and the DESeq were used for the differential expression analysis.

3. Results

3.1. Summary of MIH5 values based on entropy-based method

By using isomiRID, we mapped the next generation sequencing reads to the miRBase database and identified the location of mature sequences in reference pre-miRNAs. Based on the distribution of reads in the pre-miRNAs, we calculated the *MIH5* for each miRNA in each sample. As shown in Supplementary Table S1, we found that many mature miRNAs have 5′ variations, though most of variations show quite low frequency (mean = 0.045). The 5′ end can have substantial effects on miRNA function because these changes shift the sequence of the seed region, which is the predominant determinant of miRNA target selection.

After comparing the *MIH5* values between the early stage and the later stage of AD, we found that 47 miRNAs have significant