



In Silico Prediction and Validation of *Gfap* as an *miR-3099* Target in Mouse Brain

Shahidee Zainal Abidin^{1,2,3} · Jia-Wen Leong^{1,2,4} · Marzieh Mahmoudi^{1,5} ·
Norshariza Nordin^{2,3} · Syahril Abdullah^{2,3} · Pike-See Cheah^{1,2,4} · King-Hwa Ling^{1,2,3}

Received: 29 October 2016 / Accepted: 17 March 2017 / Published online: 8 June 2017
© Shanghai Institutes for Biological Sciences, CAS and Springer Science+Business Media Singapore 2017

Abstract MicroRNAs are small non-coding RNAs that play crucial roles in the regulation of gene expression and protein synthesis during brain development. *MiR-3099* is highly expressed throughout embryogenesis, especially in the developing central nervous system. Moreover, *miR-3099* is also expressed at a higher level in differentiating neurons *in vitro*, suggesting that it is a potential regulator during neuronal cell development. This study aimed to predict the target genes of *miR-3099* via *in-silico* analysis using four independent prediction algorithms (miRDB, miRanda, TargetScan, and DIANA-micro-T-CDS) with emphasis on target genes related to brain development and function. Based on the analysis, a total of 3,174 *miR-3099* target genes were predicted. Those predicted by at least three algorithms (324 genes) were subjected to DAVID bioinformatics analysis to understand their overall

functional themes and representation. The analysis revealed that nearly 70% of the target genes were expressed in the nervous system and a significant proportion were associated with transcriptional regulation and protein ubiquitination mechanisms. Comparison of *in situ* hybridization (ISH) expression patterns of *miR-3099* in both published and in-house-generated ISH sections with the ISH sections of target genes from the Allen Brain Atlas identified 7 target genes (*Dnmt3a*, *Gabpa*, *Gfap*, *Itga4*, *Lxn*, *Smad7*, and *Tbx18*) having expression patterns complementary to *miR-3099* in the developing and adult mouse brain samples. Of these, we validated *Gfap* as a direct downstream target of *miR-3099* using the luciferase reporter gene system. In conclusion, we report the successful prediction and validation of *Gfap* as an *miR-3099* target gene using a combination of bioinformatics resources with enrichment of annotations based on functional ontologies and a spatio-temporal expression dataset.

Electronic supplementary material The online version of this article (doi:10.1007/s12264-017-0143-0) contains supplementary material, which is available to authorized users.

✉ King-Hwa Ling
lkh@upm.edu.my

- ¹ NeuroBiology and Genetics Group, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia
- ² Genetics and Regenerative Medicine Research Center, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia
- ³ Medical Genetics Unit, Department of Biomedical Sciences, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia
- ⁴ Department of Human Anatomy, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia
- ⁵ Institute of Biological Sciences, Faculty of Science, University of Malaya, 50603 Kuala Lumpur, Malaysia

Keywords Target gene · Neurogenesis · *In silico* · Astroglialogenesis · Bioinformatics

Introduction

MicroRNAs (miRNAs) are a family of small non-coding RNAs with potent regulatory roles in metabolism, neurodevelopment, neuroplasticity, apoptosis, and other neurobiological processes [1]. MiRNAs function through partial complementary base-pairing with specific target mRNAs, resulting in the repression of translational processes or the promotion of mRNA deadenylation leading to degradation [2]. In animals, most miRNAs are partially complementary to the 3' untranslated regions (3'UTRs) of mRNA and only sporadically to 5'UTRs [3–5]. Based on

this biogenesis, bioinformaticians have created algorithms that allow prediction of the downstream targets of miRNAs. However, these bioinformatics predictions of miRNA-mRNA interactions remain challenging due to the interaction complexity and limited knowledge of the rules governing these processes. Therefore, many algorithms have been developed to predict the probability of miRNA-mRNA interactions. The computational analyses have different approaches to predicting the target genes of miRNAs, from the modelling of physical interactions to the incorporation of machine learning [6].

Most of the prediction tools use four main features to predict miRNA-mRNA target interactions: canonical site, conservation, free energy, and accessibility (secondary structure). Some prediction tools perform analysis on non-canonical sites such as G:U wobbles or mismatches in the seed region. In miRanda [7], the algorithm was developed to identify potential target genes based on the conservation, free energy, accessibility, and canonical and non-canonical site attributes. This algorithm also predicts multiple binding sites within a target gene. The support vector regression score (SVR) is calibrated based on experimental analysis of the relationship between miRNA and mRNA and interpreted as an empirical probability of target inhibition. This allows the prediction of target sites containing AU flanking content and is based on the position of the target site and UTR length [7]. Thus, it provides the strength of the miRNA regulatory effect. However, this algorithm did not incorporate the latest version of miRBase dataset for prediction [6].

Prediction programs such as TargetScan are based on the complementary binding of a seed region and categorization into three groups based on the length of exact matching (8mer, 7mer-m8, and 7mer-1A) [8]. Other parameters considered by this algorithm are the pairing contribution outside the seed region, AU content, and the distance to the nearest end of the annotated UTR of the target genes. This includes the conservation of seed regions among orthologous 3'UTRs within the miRNA binding region [9, 10]. The prediction, however, is less conservative on miRNA-mRNA interactions with wobble pairing and bulges, especially within the 5'UTR of an mRNA [11].

In DIANA-microT-CDS, the algorithm uses a minimum energy of potential miRNA binding to target genes for prediction [12], thus allowing mismatches, and the determination of a single-site prediction. The algorithm also provides a scoring and percentage probability of predicted genes being targeted in conserved and non-conserved sites [11]. In addition, DIANA-microT-CDS uses a machine-learning approach to identify the target site at the coding sequence (CDS). The algorithm predicts the accessibility of the 3'UTR by using the Sfold and AU content [12, 13].

In miRDB, the prediction of miRNA:mRNA interactions is based on both the 3'UTR and 5'UTR regions of conserved and non-conserved genes, the base composition in the regions flanking the seed pairing sites, secondary structure, and the location of the site within the 3'UTR [14]. This algorithm, however, fails to predict multiple binding sites within the target gene [6].

In this study, we used all four algorithms (miRanda, TargetScan, DIANA-microT-CDS, and miRDB) to predict and subsequently validated a downstream target of the miRNA of interest, *miR-3099*. *MiR-3099* was discovered through the deep sequencing of RNA isolated from the E (embryonic day) 15.5 developing mouse brain [15]. *MiR-3099* has a unique expression profile throughout embryogenesis. It is observed as early as the blastocyst stage, and its expression is maintained throughout embryogenesis until E11.5. Later, its expression is restricted to the cortical plate of the developing brain between E13.5 and E17.5, coinciding with the time when the majority of cells are committed to the neuronal lineage. *MiR-3099* is also more strongly expressed in retinoic acid-induced neuro-differentiating P19 teratocarcinoma cells than in proliferating and uninduced cells [15]. In a different study, *miR-3099* was found to be upregulated by 2–3-fold when 46C mouse embryonic stem cells are induced into the neural lineage [16]. Collectively, *miR-3099* has been postulated to play a role in modulating and regulating key markers involved in neuro-differentiation or neuronal cell function. Therefore, we determined the downstream targets of *miR-3099* related to brain development and function and identified and validated the glial fibrillary acidic protein (*Gfap*) gene (an important marker of astrocytes) as the primary target of *miR-3099*.

Materials and Methods

MicroRNA Target Prediction Tools

The data mining tools miRanda (<http://www.microrna.org/v3.3a>, last updated 8/2010), miRDB (<http://mirdb.org> last updated 4/2012), DIANA-microT-CDS (<http://www.microRNA.gr/microT-CDS> v5.0 last updated 7/2012), and TargetScan (<http://www.targetscan.org> v6.2 last updated 6/2012) were used to predict target genes of *miR-3099*. The predicted genes were selected according to the parameter set for each bioinformatics tool. In the miRanda algorithm, a score cut-off of 120 with a good miRSVR score was used to filter the target gene list [7]. The predicted target genes in miRDB v5.0 were downloaded using default parameters where the prediction scores were in the range of 50 to 100 [14]. DIANA-microT-CDS v5.0 was used to perform *in silico* prediction with a 0.6 score as the threshold [17]. In

TargetScan v6.2, the predicted genes were selected with a context+ score less than -0.04 [18]. Both conserved and non-conserved sites were used to cover a wide range of target sites across various species. The predicted target genes from all the tools were then illustrated using Venn Diagram v2.0 (<http://bioinfo.gp.cnb.csic.es/tools/venny/>).

Functional Annotation Clustering and Pathway Analysis

The targeted genes predicted by at least three algorithms were subjected to functional annotation clustering analysis in the Database for Annotation, Visualization and Integrated Discovery (DAVID v6.7) [19]. We performed two functional annotation analyses with DAVID. First, functional annotation clustering of selected predicted genes was performed against Gene Ontology as well as the Biocarta and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases, with medium stringency settings (kappa similarity term overlap of 3 or more, similarity threshold of 0.50, minimum 3 group members *ab initio* with EASE scores of 1 or more) to characterize the overall role and function of *miR3099* target genes. Second, the same selected list of genes was analyzed by comparison with the UP_TISSUE dataset of DAVID Knowledgebase v6.7 at a minimum threshold count of 3 and EASE scores of 0.05 or lower to determine the list of genes with expression found in nervous-system-related tissues. The shortlisted genes were subjected to expression pattern comparison as described in the following section. All statistical comparisons were made based on Fisher's exact test. All the known mouse genes were used as the background.

Comparison of Expression Profiles of *miR-3099* and Its Target Genes Using the Allen Mouse Brain and Developing Mouse Brain Atlases

Gene expression profiles of all the predicted target genes by at least three databases were screened and manually interpreted based on the Allen Mouse Brain and Developing Mouse Brain Atlases (<http://developingmouse.brain-map.org/>). First, ISH profiles of *miR-3099* on E11.5-to-E17.5 brain sections published in Ling *et al.* (2011) [15] were compared with the ISH expression of the predicted target genes on the corresponding E11.5-to-E18.5 sections in the Allen Mouse Developing Brain Atlas. In principle, miRNA is involved in post-transcriptional gene silencing through complementary base-pairing with mRNA. Thus, we proposed the expression profiles of predicted target genes would be suppressed between E11.5 until E18.5, especially in the telencephalon. The gene expression was summarized in heatmap format with the age of the mouse on the vertical axis and the anatomic region on the horizontal axis.

Tissue Collection

C57BL/6 mice were maintained under a 12 h light/12 h dark cycle with free access to food and water. No pharmacological treatments were given. All mouse studies were approved by the Animal Care and Use Committee of Universiti Putra Malaysia (UPM/FPSK/PADS/BR-UUH/00469). Transcardial perfusion was performed on adult mice P (postnatal day) 86 prior to brain harvesting. The mice were anesthetized with sodium pentobarbital (50 mg/kg body weight) *via* intraperitoneal injection. Upon absence of the toe-pinch withdrawal reflex, the thorax was dissected to expose the heart. A 25G needle was inserted to the apex and the right atrium was lacerated. The mouse was first perfused with 1× phosphate buffered saline (PBS) followed by 4% paraformaldehyde (PFA), pH 7.0. When perfusion was completed, the mouse was decapitated and the brain was harvested. Newborn pups (P1.5) were not perfused but rapidly decapitated by using a pair of sharp surgical scissors for immediate brain harvesting.

Locked Nucleic Acid-*In Situ* Hybridization

Mouse brains (P1.5 and P86) were harvested and fixed with 4% PFA in 1× PBS, pH 7.0 for 2 days followed by several washes with 70% ethanol. The tissue was subjected to a standard tissue processing protocol in which it was dehydrated in a series of ethanol washes and xylene using a Leica TP1020 (Nussloch, Germany) semi-closed bench-top tissue processor. The tissue was embedded in paraffin wax and sectioned at 8 µm for locked nucleic acid-*in situ* hybridization (LNA-ISH) with a protocol adapted from a previous study [15]. The sections were deparaffinized with three washes in xylene (5 min each), hydrated in a series of ethanol washes with RNase-free water, and then fixed in 4% PFA in 1× PBS for 10 min followed by proteinase K digestion [1.2 µg/µL proteinase K, 0.1 mol/L Tris-HCl (pH 7.5), 0.05 mmol/L EDTA] for 30 min. Then the sections were re-fixed in 4% PFA in 1× PBS for 5 min and acetylated [1.3% triethanolamine, 0.178% concentrated HCl, 0.25% acetic anhydride] for 10 min. Sections were washed with 1× PBS between each step.

In the pre-hybridization step, the sections were incubated at 65 °C for 2 h in pre-hybridization buffer that consisted of 50% deionized formamide, 3× sodium chloride/sodium citrate (SSC), 1× Denhardt's solution, 10% Dextran sulfate, 1 mg/mL yeast total RNA, 1 mg/mL Herring Sperm DNA, and 1× PBS. Then, *miR-3099* LNA probe (Cat. no: 99999-15, Exiqon, Vedbaek, Denmark) was added to the buffer to give a concentration of 0.02 pmol/µL while *miR-3099** LNA probe was used as a negative control for hybridization. Hybridization was carried out in

a hybridization chamber at 60 °C overnight. Next, the sections were washed in serial concentrations of SSC (2×, 1×, 0.5× and 0.1×) for 15 min at 48 °C, and then briefly rinsed in fresh 0.1× SSC followed by incubation in pre-blocking buffer [0.1 mol/L Tris-HCl (pH 7.5), 0.15 mol/L NaCl, 240 µg/mL levamisole] for 5 min at room temperature.

In a humidified chamber, the sections were blocked in 5% heat-inactivated fetal calf serum, 0.1% Tween-20, and 1% blocking powder in maleate buffer for 1 h at room temperature. Then, the sections were incubated with anti-digoxigenin antibody with alkaline phosphatase conjugate, Fab fragments (Roche Diagnostics, Branford, CT) in blocking buffer (1:1000) for 1 h in the dark. Subsequently, the sections were washed three times (10 min each) in pre-blocking buffer at room temperature. The sections were then washed with NTMT buffer (0.1 mol/L Tris-HCl (pH 9.5), 0.1 mol/L NaCl, 0.05 mol/L MgCl₂, 15% Tween-20 and 240 µg/ml levamisole) for a further 10 min at room temperature. Then, the sections were incubated with nitro blue tetrazolium chloride/5-bromo-4-chloro-3-indolyl phosphate, toluidine salt in 0.1 mol/L Tris-HCl (pH 9.5), 0.1 mol/L NaCl, 0.05 mol/L MgCl₂, 15% Tween-20, and 1× levamisole for 48 h. After the color reaction step, the sections were washed with Tris-EDTA buffer pH 8.0 (containing 0.01 mol/L Tris-HCl, pH 7.5, and 0.001 mol/L EDTA, pH 8.0) for 10 min and dehydrated in a series of ethanol concentrations and xylene for 3 min between each step. The sections were mounted with DPX mounting medium prior to observation under a Nikon SMZ745T (New York, NY) stereomicroscope.

Dual Luciferase Assay

The downstream targets of *miR-3099* were examined by measuring luciferase activity using a luciferase reporter containing the glial fibrillary acidic protein (*Gfap*) 3'UTR. The UTR region of *Gfap* (NM_001131020.1) and the *miR-3099* precursor sequence were amplified from mouse cDNA and genomic DNA, respectively, using the following primers: UTR region of *Gfap* 5'GTC CTT TGT CTA GAA TTC CAG GCC CA'3 (sense) and 5'AGG CAT TTG CTC GAG GGT TCA GAG TAT'3 (anti-sense), *miR-3099* 5' CTT CCT GTG CTT TCT GAG ATC TTG TAA G'3 (sense) and 5'GTG TTT TCC TGC GAA TTC TCA TCA GAT G'3 (anti-sense). The polymerase chain reaction (PCR) was performed in an Eppendorf Mastercycler® (Hamburg, Germany) with a pre-denaturation step at 94 °C for 2 min, followed by 40 cycles of 94 °C for 15 s, 59 °C for 15 s, and 72 °C for 2 min with an additional elongation at 72 °C for 5 min. The PCR was carried out using Expand Long Template PCR System Kit (Roche Diagnostics) with 300 nmol/L sense and anti-sense primers, and 50 ng cDNA

or genomic DNA. The amplified *Gfap* UTR amplicon was digested with *EcoRI* and *XhoI* before ligating the insert into pEZX-MT01 plasmid containing the luciferase reporter gene (GeneCopoeia, Rockville, MD). On the other hand, *miR-3099* was digested with *BglIII* and *EcoRI* and cloned into pEZX plasmid that carried eGFP (GeneCopoeia).

For transfection, the human embryonic kidney (HEK) 293FT cell line was used and maintained in Dulbecco's modified Eagle's medium (Gibco, Massachusetts) supplemented with 2 mmol/L glutamine (Gibco), 100 µg/mL penicillin/streptomycin (Invitrogen, Massachusetts), and 10% (v/v) fetal bovine serum (Gibco). Transfection was performed using Lipofectamine 3000 (Invitrogen) in 6-well plates during the suspension stage at 8.0×10^5 cells/well. The HEK293FT cells were transfected with 2 µg pEZX-*Gfap* alone (group 1), or co-transfected with 2 µg pEZX-*Gfap* and 2 µg pEZX-*miR-3099* (group 2), 2 µg empty plasmid and 2 µg pEZX-*miR-3099* (group 3), and 2 µg pEZX-*Gfap* and 2 µg pEZX-miR-scrambled (group 4). The transfection procedures and luciferase assays were performed according to the manufacturer's instructions. After 48 h of transfection, the cells were washed with 1× PBS prior to incubation with 1× lysis buffer for 10 min on an orbital shaker at room temperature. Firefly and *Renilla* luciferase (GeneCopoeia) activities were measured using an Asys UVM30 (Cambridge, UK) microplate reader. Data were generated after normalization to *Renilla* luciferase and expressed as the relative luciferase activity. The experiments were repeated in three biological replicates.

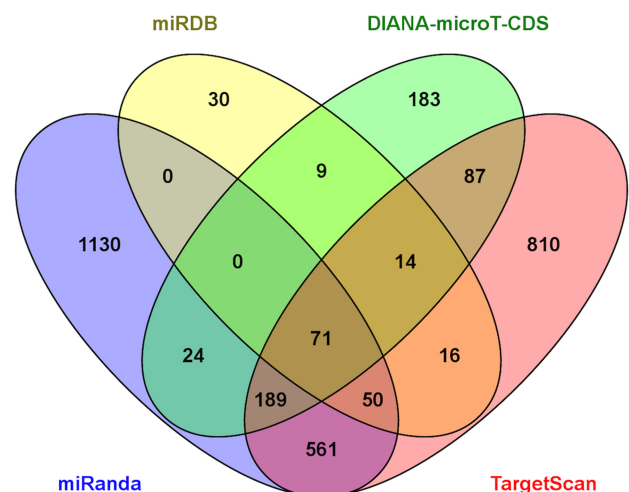


Fig. 1 Venn diagram depicting the number of predicted target genes of *miR-3099* from miRanda, miRDB, TargetScan, and DIANA-microT-CDS. The overlaps indicate the numbers of genes predicted by more than one algorithm. miRanda, miRDB, TargetScan, and DIANA-microT-CDS predicted a total of 2025, 190, 1798, and 577 genes, respectively.

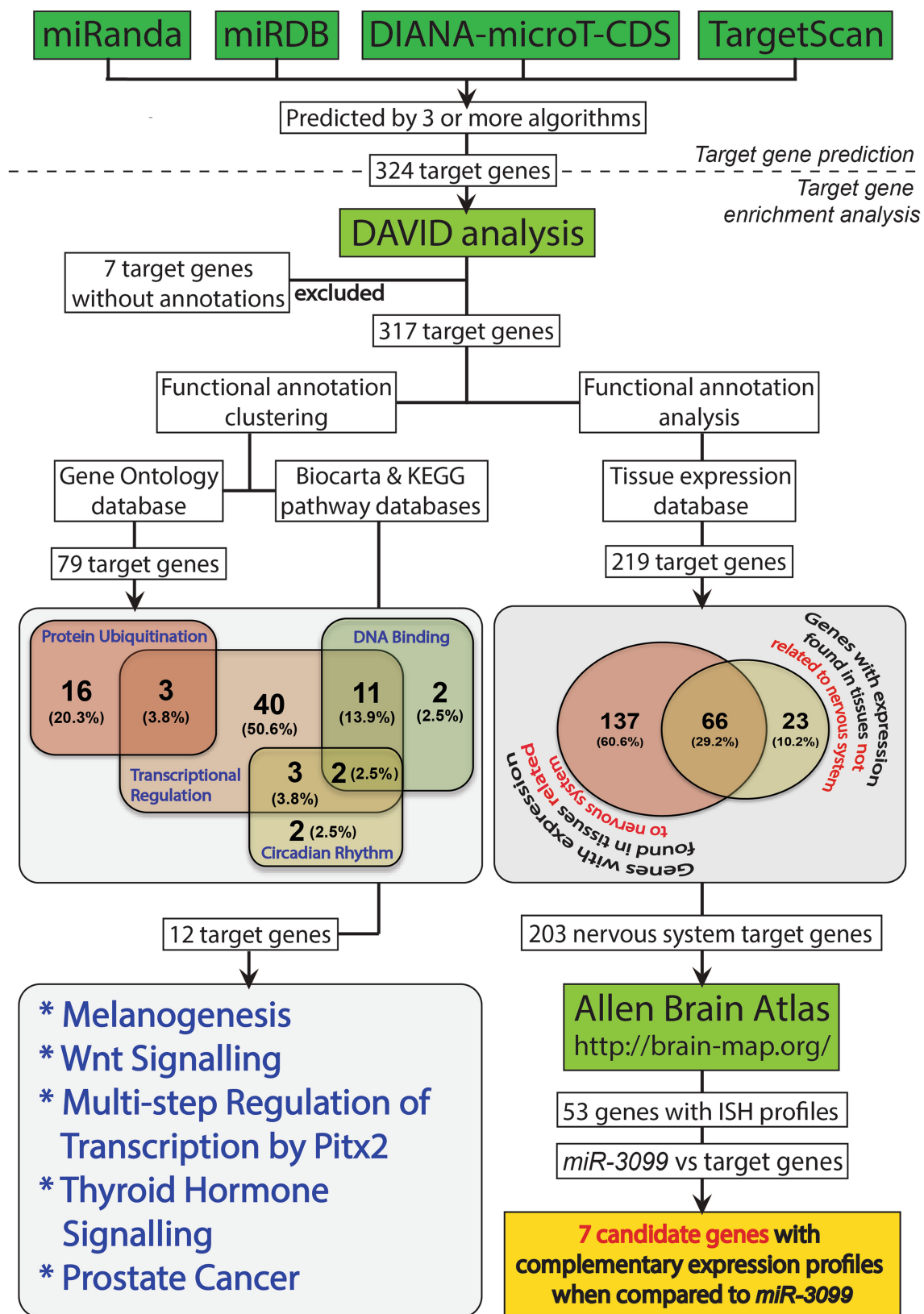


Fig. 2 Pipeline for *miR-3099* target gene prediction and functional characterization based on DAVID analysis (based on Gene Ontology, Biocarta, KEGG and tissue expression databases) and comparison with the Allen Brain Atlas *in situ* hybridization profile.

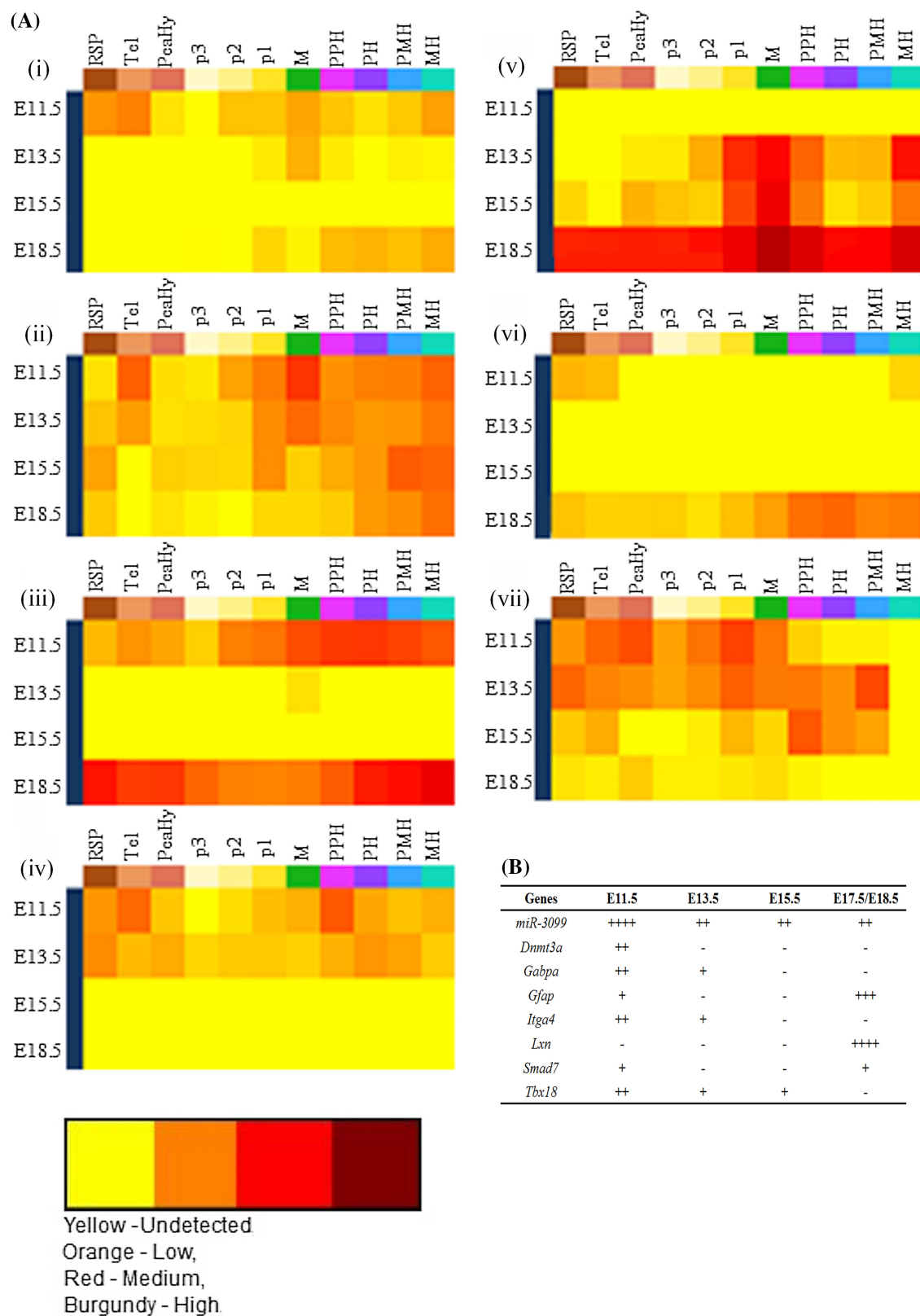


Fig. 3 Heatmap of the expression energy of 8 genes across 11 brain regions during E11.5–E18.5 and the expression pattern of *miR-3099* from E11.5 to E17.5. **A** Heatmaps for (i) *Dnmt3a*, (ii) *Gabpa*, (iii) *Gfap*, (iv) *Itga4*, (v) *Lxn*, (vi) *Smad7*, and (vii) *Tbx18* were retrieved from the Allen Developing Mouse Brain Atlas. Each heatmap illustrates increasing expression energy (expression level) ranging from a log value of -1.5 in yellow to 3.5 in burgundy. RSP, rostral secondary prosencephalon; Tel, telencephalic vesicle; PedHy, peduncular hypothalamus; p3, prosomere 3; p2, prosomere 2; p1, prosomere 1; M, midbrain; PPH, prepontine hindbrain; PH, pontine hindbrain; PMH, pontomedullary hindbrain; MH, medullary hindbrain. **B** Summary of expression profile of *miR-3099* (E11.5–E17.5) and the predicted target genes (E11.5–E18.5) in the telencephalon. –, undetected (yellow); + and ++, low expression (orange); +++ and +++++, moderate expression (red); ++++++ and ++++++, high expression (burgundy).

Statistical Analysis

One-way ANOVA and *post-hoc* Bonferroni tests were used to compare the firefly expression levels among the transfected groups. Differences between control and treatment groups were considered significant at $P < 0.05$.

Results

In Silico Predicted Downstream Targets of *miR-3099*

The prediction of target genes was based on the currently-known mammalian sequence and scanning of the 3'UTRs of the mouse (*Mus musculus*) genome for potential target sites. The computational analysis of *miR-3099* target genes based on the four data mining tools identified a total of 3,174 genes (both conserved and non-conserved miRNA targets) (Table S1). Of these, 324 were predicted by at least three of the algorithms (Fig. 1). The combination of multiple algorithms provides more accurate predicted target genes and reduces false-positives [20].

Functional Characterization of *miR-3099* Target Genes

To characterize the overall roles and functions of the predicted target genes of *miR-3099*, DAVID analysis was performed to analyze 324 genes (predicted by at least three different algorithms) for functional annotation clustering analysis. Of these genes, 317 (Supplementary Table S2) with associated annotation terms were subjected to functional annotation clustering based on the Gene Ontology database and then the Biocarta and KEGG pathways (Fig. 2). A total of 11 annotation clusters were significantly represented by the genes, but only 4 had an EASE score of 1.0 or more. These clusters were represented by 79 genes (non-redundant) and related to molecular functions or

biological processes involved in protein ubiquitination (19 genes), DNA binding (16 genes), transcriptional regulation (59 genes), and circadian rhythm (7 genes) (Table S3). Using the Biocarta and KEGG databases, only one cluster with an EASE score of 1 was identified. A total of 12 genes were clustered among pathways associated with melanogenesis, *Wnt* signaling, multi-step regulation of transcription by *Pitx2*, thyroid hormone signaling, and cancers such as prostate cancer (Table S4).

Shortlist of Candidate Target Genes

To select candidates involved in brain development and function, the genes were shortlisted based on the localization of expression in the nervous system. From analysis against the UP_TISSUE dataset (Table S5), 307 out of 317 genes were annotated using the DAVID knowledgebase with 219 non-redundant genes significantly mapped to one or more of 13 different tissues, of which 8 were nervous system-related (brain, cerebellum, fetal brain, brain cortex, neural stem cells, olfactory bulb, spinal cord, and spinal ganglion) (203 genes; Table S6). A total of 89 genes were mapped to non-nervous system tissues such as skin, pancreatic islet, eye, mesenchymal stem cell, and lung (Fig. 2).

The expression profile for each of the 203 targeted genes was explored using the Allen Mouse Brain and Developing Mouse Brain Atlases, and only 53 were available. Only 7 of the 53 genes (Fig. 3) had expression profiles complementary to the *miR-3099* expression profile described previously [15]. These genes were DNA (cytosine-5-)methyltransferase 3 alpha (*Dnmt3a*), *Gfap*, GA binding protein transcription factor alpha subunit (*Gabpa*), integrin alpha 4 (*Itga4*), latexin (*Lxn*), SMAD family 7 (*Smad7*), and T-box 18 (*Tbx18*). Most of them had low-to-undetectable expression in the telencephalon except for *Gfap* and *Lxn*, which had moderate expression at E18.5. On the other hand, *miR-3099* was moderately expressed in the telencephalon at E11.5, while it was expressed at a low level between E13.5 and E17.5 (Fig. 3) [15]. To further validate the complementary expression profile of *miR-3099* with its target genes in the adult brain, we compared our pre-existing ISH of *miR-3099* sections generated at post-natal and adult time-points (P1.5 and P86) to the sections from the P4 and P56 mouse brain in the Allen Mouse Brain Atlas. *miR-3099* was highly expressed in the cerebral cortex (CC) and the hippocampus (HPC) at both P1.5 and P86 (Fig. S1). Assuming insignificant anatomical, functional, and molecular differences between P1.5 and P4 brains, the P1.5 *miR-3099* versus P4 target gene comparison indicated 3 genes (*Dnmt3a*, *Gfap*, and *Lxn*) that showed low expression in the CC and moderate expression in the HPC. In the P86 *miR-3099* versus P56 target gene

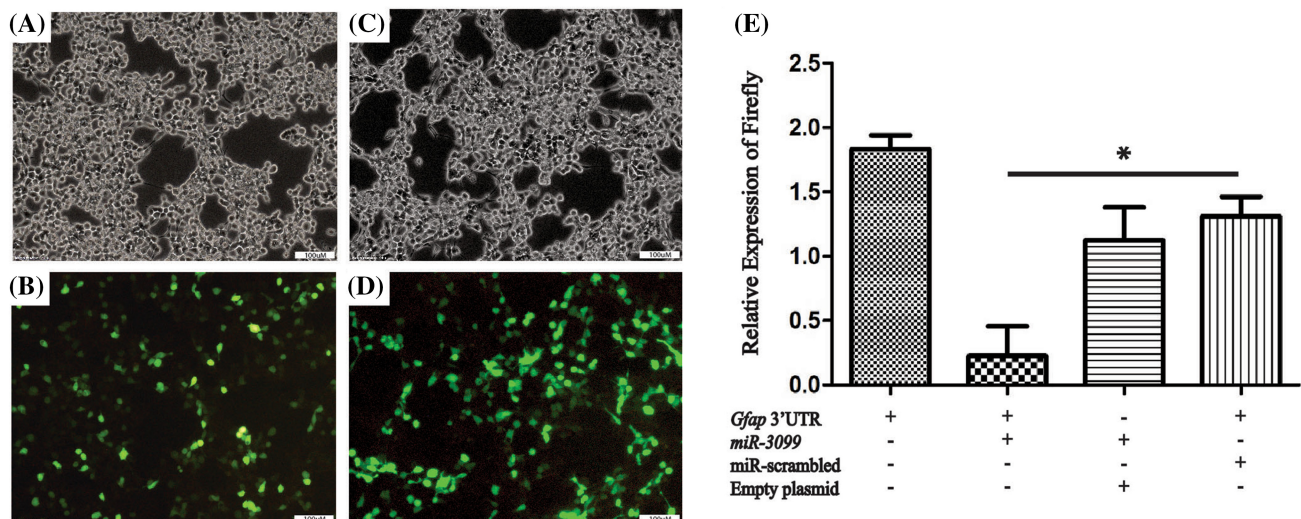


Fig. 4 Phase-contrast (A, C) and fluorescence (B, D) micrographs of HEK293FT cells transfected with miR-scramble (A, B) and *miR-3099* (C, D). **E** Relative luciferase activity of *Gfap* 3'UTR measured after 48 h of co-transfection with either *miR-3099* or miR-scrambled vector, which demonstrated significant down-regulation of luciferase

due to *miR-3099* but not miR-scrambled expression. Luciferase levels are reported as fold-change compared to the values measured in control (miR-scrambled co-transfected with *Gfap*) after normalization to *Renilla* luciferase activity (* $P < 0.05$; ANOVA with Bonferroni *post-hoc* test).

comparison, the expression of *Smad7*, *Dnmt3a*, *Gfap*, and *Lxn* was found to be low or moderate in the CC. In addition, the *Dnmt3a* and *Gfap* genes were expressed at a low level in the HPC, suggesting that *miR-3099* is a potential regulator of these genes (Figs. S2 and S3).

Validation of Predicted Target Gene via Luciferase Assay

Based on the prediction of downstream target analysis and expression profile in the Allen Brain Atlas, *Gfap* was selected for further validation using luciferase gene reporter assays. GFAP is a regulatory factor in astroglial fate determination during brain development through the JAK-STAT and Notch signaling pathways [21, 22]. The 3'UTR of *Gfap* was cloned downstream to the luciferase coding sequence in the PEZX-MT01 vector. The level of luciferase activity was then measured as a function of *miR-3099* in targeting the cloned *Gfap* 3'UTR. The luciferase activity was inversely correlated with the specificity of *miR-3099* in targeting *Gfap*. The transfection efficiency assessed by GFP expression from the miRNA expression was ~80% for both *miR-3099* and miR-scrambled (Fig. 4). Four different groups were involved in the study. Group 1 served as a negative control that determines the effect of host intrinsic factors on the chimeric target. Group 2 was the assay test group. Groups 3 and 4 were negative controls that determined the specificity of *miR-3099* to the target gene. Our results demonstrated that *Gfap* was targeted by *miR-3099* (Group 2) and showed a reduced

luciferase signal ($P < 0.05$) compared to Group 4. HEK293FT cells co-transfected with miR-scrambled and pEZX-MT01-*Gfap* vector (Group 4) showed no changes in the levels of luciferase signals, indicating that the 3'UTR of *Gfap* is targeted specifically by *miR-3099* (Fig. 4).

Discussion

The identification of *miR-3099* target genes revealed the possible role of this miRNA. Each prediction algorithm, however, has a limitation due to the difficulty of associating the predicted target genes with their potentially enriched molecular and biological functions. Some of the algorithms rely on the seed match for target prediction. This approach would exclude non-conserved or/and thermodynamically favorable target sites. Therefore, a combination of predictive target tools was used to minimize false-positively predicted target genes and provide more accurate target prediction [20]. When coupled with functional annotation analysis and comparative expression profiling of the target genes, targets of interest can be confidently identified for experimental validation.

DAVID functional annotation analysis based on the Gene Ontology, Biocarta/KEGG, and tissue expression databases showed that a large proportion of *miR-3099* target genes occurred in tissues and pathways associated with the nervous system. The analysis was in agreement with the expression pattern of *miR-3099*, which was found predominantly in the developing brain [15], indicating that

the downstream targets have functions related to nervous system development. Interestingly, the *Wnt* signaling pathway and multi-step regulation of *Pitx2* gene pathway found here are known to be involved in self-renewal and neural differentiation processes [21, 23, 24]. Therefore, *miR-3099* may play a role in modulating and regulating key markers involved in these pathways.

Comparative ISH analysis using published embryonic brain sections [15] and in-house-generated postnatal and adult brain sections provided additional filters for selecting relevant target genes for downstream validation. The ISH dataset used in comparative analysis at the postnatal and adult time-points did not directly match the ISH dataset from the Allen Brain Atlas. Although there were no significant anatomical and functional differences between P1.5/P4 and P86/P56, there may have been underlying molecular changes. The different ISH methods used in this study and the Allen Brain Atlas contributed more variables to the qualitative comparative analysis. The analysis should be carefully interpreted and used as supplementary evidence to support the major findings based on the embryonic ISH comparison.

Gfap was chosen as a candidate for downstream validation because it had been implicated in an important pathway in the brain known as the JAK-STAT signaling pathway, which controls the onset of astrocyte formation [21]. During astrocyte formation, STAT3 induces SMAD family member 1 (*Smad1*) activation to form the STAT3-Ep300-Smad complex. STAT3 also mediates BMP2 expression and subsequently activates *Smad1* [25], which promotes astrogliogenesis *via* *Gfap* expression. Using reporter assays, we demonstrated that *miR-3099* targeted the 3'UTR of *Gfap*. This finding is in line with a recent study by Choi *et al.* [26]. Interestingly, the *Gfap* knockout in astrocytes favors neuronal survival and supports neurite growth [27]. Moreover, overexpression of laminin and punctate aggregation of fibronectin occur in the absence of GFAP [27]. These components are important in the migratory pathway of neural crest cells [28, 29].

Besides *miR-3099*, other miRNAs such as *miR-330*, *miR-326*, and *miR-145* negatively regulate *Gfap* [26, 30]. This indicates that *Gfap* has multiple target sites for various miRNAs and forms a complex regulatory network involving miRNA-mRNA interactions. Thus, the interactions are not based on a one-to-one relationship but a cluster of miRNAs that co-operatively regulate *Gfap* in astroglial fate determination. To allow a more robust analysis of complex miRNA-mRNA-protein relationships such as one-to-many, many-to-one, and many-to-many, high-throughput transcriptomic and proteomic analyses should be performed and analyzed together. Combinations of approaches such as qPCR, microarray, bioinformatics analysis, and quantitative liquid chromatography coupled with tandem mass spectrometry

have been simultaneously used to determine the many-to-many relationships of miRNA, mRNA, and translational repression in Chinese hamster ovary cells [31]. Next-generation sequencing of small RNAs and RNA-seq and pathway analysis have been used to profile the relationships of dysregulated miRNAs with mRNAs in primary posterior longitudinal ligament cells isolated from patients with ossification of the posterior longitudinal ligament and normal patients [32]. Simultaneous data generation, bioinformatics analysis, and interpretation of miRNA, mRNA, and proteomics datasets from comparative studies will provide better understanding of the complex regulatory networks regulated by miRNAs.

In summary, *Gfap* was predicted and validated as one of the downstream targets of *miR-3099*. Using multiple miRNA target prediction tools and cross-comparison with their functional annotation analysis and comparative expression profiling provides more confident insights and experimental targets of novel miRNAs, thus helping to reduce the need to systematically validate all or large numbers of target genes *via* experimental means.

Acknowledgements This work was supported by the Science Fund (02-01-04-SF2336) and the Fundamental Research Grant Scheme, Ministry of Higher Education, Malaysia (FRGS-04-01-15-1663FR). ZAS was a recipient of a Malaysian Ministry of Higher Education MyBrain15 scholarship.

References

1. Kosik KS, Krichevsky AM. The elegance of the microRNAs: A neuronal perspective. *Neuron* 2005, 47: 779–782.
2. Ambros V. The functions of animal microRNAs. *Nature* 2004, 431: 350–355.
3. Lytle JR, Yario TA, Steitz JA. Target mRNAs are repressed as efficiently by microRNA-binding sites in the 5' UTR as in the 3' UTR. *Proc Natl Acad Sci U S A* 2007, 104: 9667–9672.
4. Tay Y, Zhang J, Thomson AM, Lim B, Rigoutsos I. MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation. *Nature* 2008, 455: 1124–1128.
5. Forman JJ, Legesse-Miller A, Collier HA. A search for conserved sequences in coding regions reveals that the let-7 microRNA targets Dicer within its coding sequence. *Proc Natl Acad Sci U S A* 2008, 105: 14879–14884.
6. Peterson SM, Thompson JA, Ufkin ML, Sathyanarayana P, Liaw L, Congdon CB. Common features of microRNA target prediction tools. *Front Genet* 2014, 5: 23. doi: 10.3389/fgene.2014.00023.
7. Betel D, Koppal A, Agius P, Sander C, Leslie C. Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. *Genome Biol* 2010, 11: R90.
8. Grimson A, Farh KK-H, Johnston WK, Garrett-Engle P, Lim LP, Bartel DP. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol Cell* 2007, 27: 91–105.
9. Friedman RC, Farh KK-H, Burge CB, Bartel DP. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res* 2009, 19: 92–105.

10. Nielsen CB, Shomron N, Sandberg R, Hornstein E, Kitzman J, Burge CB. Determinants of targeting by endogenous and exogenous microRNAs and siRNAs. *RNA* 2007, 13: 1894–1910.
11. Witkos TM, Koscińska E, Krzyżosiak WJ. Practical aspects of microRNA target prediction. *Curr Mol Med* 2011, 11: 93–109.
12. Paraskevopoulou MD, Georgakilas G, Kostoulas N, Vlachos IS, Vergoulis T, Reczko M, *et al.* DIANA-microT web server v5.0: service integration into miRNA functional analysis workflows. *Nucleic Acids Res* 2013, 41: W169–173.
13. Chan CY, Lawrence CE, Ding Y. Structure clustering features on the Sfold Web server. *Bioinformatics* 2005, 21: 3926–3928.
14. Wong N, Wang X. miRDB: an online resource for microRNA target prediction and functional annotations. *Nucleic Acids Res* 2015, 43: D146–152.
15. Ling KH, Brautigan PJ, Hahn CN, Daish T, Rayner JR, Cheah PS, *et al.* Deep sequencing analysis of the developing mouse brain reveals a novel microRNA. *BMC Genomics* 2011, 12: 176.
16. Zainal Abidin S, Abbaspourbabaei M, Ntimi CM, Siew WH, Pike-See C, Rosli R, *et al.* MiR-3099 is overexpressed in differentiating 46c mouse embryonic stem cells upon neural induction. *Malays J Med Sci* 2014, 21: 27–33.
17. Vlachos IS, Kostoulas N, Vergoulis T, Georgakilas G, Reczko M, Maragkakis M, *et al.* DIANA miRPath v2.0: investigating the combinatorial effect of microRNAs in pathways. *Nucleic Acids Res* 2012, 40: W498–504.
18. Garcia DM, Baek D, Shin C, Bell GW, Grimson A, Bartel DP. Weak seed-pairing stability and high target-site abundance decrease the proficiency of lsi-6 and other microRNAs. *Nat Struct Mol Biol* 2011, 18: 1139–1146.
19. Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* 2009, 37: 1–13.
20. Zhang Y, Verbeek FJ. Comparison and integration of target prediction algorithms for microRNA studies. *J Integr Bioinform* 2010, 7: 127.
21. Wen S, Li H, Liu J. Dynamic signaling for neural stem cell fate determination. *Cell Adh Migr* 2009, 3: 107–117.
22. Chaboub LS, Deneen B. Astrocyte form and function in the developing central nervous system. *Semin Pediatr Neurol* 2013, 20: 230–235.
23. Hao J, Li TG, Qi X, Zhao D-F, Zhao GQ. WNT/beta-catenin pathway up-regulates Stat3 and converges on LIF to prevent differentiation of mouse embryonic stem cells. *Dev Biol* 2006, 290: 81–91.
24. Cheffer A, Tárnok A, Ulrich H. Cell cycle regulation during neurogenesis in the embryonic and adult brain. *Stem Cell Rev* 2013, 9: 794–805.
25. Fukuda S, Abematsu M, Mori H, Yanagisawa M, Kagawa T, Nakashima K, *et al.* Potentiation of astrogliogenesis by STAT3-mediated activation of bone morphogenetic protein-Smad signaling in neural stem cells. *Mol Cell Biol* 2007, 27: 4931–4937.
26. Choi I, Woo JH, Jou I, Joe EH. PINK1 deficiency decreases expression levels of mir-326, mir-330, and mir-3099 during brain development and neural stem cell differentiation. *Exp Neurobiol* 2016, 25: 14–23.
27. Menet V, Giménez y Ribotta M, Chauvet N, Drian MJ, Lannoy J, Colucci-Guyon E, *et al.* Inactivation of the glial fibrillary acidic protein gene, but not that of vimentin, improves neuronal survival and neurite growth by modifying adhesion molecule expression. *J Neurosci* 2001, 21: 6147–6158.
28. Brauer PR, Markwald RR. Specific configurations of fibronectin-containing particles correlate with pathways taken by neural crest cells at two axial levels. *Anat Rec* 1988, 222: 69–82.
29. Sheppard AM, Hamilton SK, Pearlman AL. Changes in the distribution of extracellular matrix components accompany early morphogenetic events of mammalian cortical development. *J Neurosci* 1991, 11: 3928–3942.
30. Wang CY, Yang SH, Tzeng SF. MicroRNA-145 as one negative regulator of astrogliosis. *Glia* 2015, 63: 194–205.
31. Clarke C, Henry M, Doolan P, Kelly S, Aherne S, Sanchez N, *et al.* Integrated miRNA, mRNA and protein expression analysis reveals the role of post-transcriptional regulation in controlling CHO cell growth rate. *BMC Genomics* 2012, 13: 656.
32. Xu C, Chen Y, Zhang H, Chen Y, Shen X, Shi C, *et al.* Integrated microRNA-mRNA analyses reveal OPLL specific microRNA regulatory network using high-throughput sequencing. *Sci Rep* 2016, 6: 21580.