

RESEARCH

Open Access



# Functional enrichment analysis based on long noncoding RNA associations

Kuo-Sheng Hung<sup>1</sup>, Chung-Chi Hsiao<sup>1</sup>, Tun-Wen Pai<sup>1\*</sup>, Chin-Hwa Hu<sup>2,5</sup>, Wen-Shyong Tzou<sup>2,5</sup>, Wen-Der Wang<sup>3</sup> and Yet-Ran Chen<sup>4</sup>

From The 11th International Conference on Systems Biology (ISB 2017) Shenzhen, China. 18-21 August 2017

## Abstract

**Background:** Differential gene expression analysis using RNA-seq data is a popular approach for discovering specific regulation mechanisms under certain environmental settings. Both gene ontology (GO) and KEGG pathway enrichment analysis are major processes for investigating gene groups that participate in common biological responses or possess related functions. However, traditional approaches based on differentially expressed genes only detect a few significant GO terms and pathways, which are frequently insufficient to explain all-inclusive gene regulation mechanisms.

**Methods:** Transcriptomes of survivin (*birc5*) gene knock-down experimental and wild-type control zebrafish embryos were sequenced and assembled, and a differential expression (DE) gene list was obtained for traditional functional enrichment analysis. In addition to including DE genes with significant fold-change levels, we considered additional associated genes near or overlapped with differentially expressed long noncoding RNAs (DE lncRNAs), which may directly or indirectly activate or inhibit target genes and play important roles in regulation networks. Both the original DE gene list and the additional DE lncRNA-associated genes were combined to perform a comprehensive overrepresentation analysis.

**Results:** In this study, a total of 638 DE genes and 616 DE lncRNA-associated genes (lncGenes) were leveraged simultaneously in searching for significant GO terms and KEGG pathways. Compared to the traditional approach of only using a differential expression gene list, the proposed method of employing DE lncRNA-associated genes identified several additional important GO terms and KEGG pathways. In GO enrichment analysis, 60% more GO terms were obtained, and several neuron development functional terms were retrieved as complete annotations. We also observed that additional important pathways such as the FoxO and MAPK signaling pathways were retrieved, which were shown in previous reports to play important roles in apoptosis and neuron development functions regulated by the survivin gene.

**Conclusions:** We demonstrated that incorporating genes near or overlapped with DE lncRNAs into the DE gene list outperformed the traditional enrichment analysis method for effective biological functional interpretations. These hidden interactions between lncRNAs and target genes could facilitate more comprehensive analyses.

**Keywords:** lncRNA, Differential expression, Gene ontology (GO), KEGG, Neuron development

\* Correspondence: [twp@mail.ntou.edu.tw](mailto:twp@mail.ntou.edu.tw)

<sup>1</sup>Department of Computer Science and Engineering, National Taiwan Ocean University, Keelung, Taiwan

Full list of author information is available at the end of the article



## Background

Embryonic development of the nervous system mainly relies on complex interactions between extrinsic signaling factors and intrinsic regulation of gene expression. Different environmental conditions, such as temperature, diet, toxin levels, and chemical levels can change the gene expression profiles from their usual patterns and lead to defects in neuron development. Recent research has indicated that hypoxic (~1% O<sub>2</sub>) environments also affect neuron-related gene expression [1]. Evidence has shown that upregulated expression of hypoxia-inducible factor (HIF) in response to the absence of oxygen leads to a series of downregulated gene expressions and induces physiological and metabolic changes in neuron development [2].

Our previous studies showed that zebrafish under hypoxic embryogenesis conditions exhibited upregulated expression of *hif2a*, which is involved in neural differentiation processes via expression regulation of the survivin family (*birc5* gene) [3]. Survivin is a protein containing a BIR (Baculovirus Inhibitor of apoptosis Repeat) domain and an  $\alpha$ -helical coiled-coil domain that is expressed in neural precursor cells and neural progenitor cells during development. Survivin, as an IAP (inhibitor of apoptosis proteins) family member that binds to caspases, can inhibit apoptosis [4, 5]. Further investigation into the role of survivin-related genes in neuron development is important to the understanding of the basic mechanisms controlling neural cell growth and signal transduction pathways.

Noncoding RNAs (ncRNAs) are abundant in cells, are not translated into proteins, and are functionally categorized into two major types. The first type is structural noncoding RNA, which includes transfer RNAs, ribosomal RNAs, small nucleolar RNAs, and small nuclear RNAs; the second type is regulatory noncoding RNA, which includes microRNAs, small interfering RNAs, extracellular RNAs, piwi-interacting RNAs, and long noncoding RNAs, among others. Several studies have shown that microRNAs (miRNAs) and long noncoding RNAs (lncRNAs) possess gene expression regulation properties in neuronal system development [3, 6, 7]. The regulatory relationships between miRNA and genes have already attracted attention and have been verified by satisfactory experiments [8]. Comparatively few lncRNAs have been studied in animal models, and a comprehensive functional analysis of most lncRNAs remains challenging because lncRNAs are not evolutionarily conserved among related species.

Several kinds of lncRNAs are categorized based on their genomic loci. Genomic locations of lncRNAs associated with related protein-coding genes can be grouped as sense or antisense transcript overlapping and as promoter-associated, intronic, or intergenic. A promoter-associated lncRNA is transcribed from the upstream region of the protein-coding genes [9], an intronic lncRNA possesses a

transcribed region overlapped with the intron segments of protein-coding genes, and an intergenic lncRNA is located between two protein-coding genes [10]. Additionally, some special lncRNAs possess coding ability and form as regular peptides, which probably perform essential functions [11, 12]. Many lncRNAs were systematically identified by Pauli et al. through a time-series of RNA-seq experiments related to zebrafish embryo development. The results demonstrated that lncRNAs are expressed significantly at early embryonic stages and possess stage-specific and tissue-specific expression characteristics [13].

One traditional approach for understanding regulation mechanisms is analyzing differentially expressed genes (DEGs) through Gene Ontology (GO) terms and pathway (such as KEGG pathways) enrichment analysis using hypergeometric distributions. Unfortunately, sometimes only a few significant GO terms and pathways can be identified using the limited number of DEGs, which leads to difficulty in explaining all-inclusive gene regulation mechanisms.

Normally, the regulation of ncRNA expression is not considered to be within the overall RNA regulation network. However, many genes that undergo no significant changes in expression are regulated mainly through induction or inhibition by ncRNAs; thus, discounting the regulatory function of ncRNAs may lead to inadequate identification of significant GO terms or functional pathways. An alternative solution is to incorporate lncRNAs in regulation analysis and to apply co-expression network methods for a comprehensive analysis by using tools such as ncFANs [14], lncRNAtor [15], and Co-lncRNA [16]. Although co-expression network analysis provides possible data to bridge gaps in functional annotations, it has some limitations. Serin et al. indicated that large data connections would make co-expression networks complicated and explaining biological functionality difficult. Additionally, the connection between genes in a co-expression network does not necessarily imply the existence of a regulatory relationship in biological systems [17].

In this study, we proposed a method of combining genes near or overlapped with differentially expressed lncRNAs (DE lncRNAs) and DEG lists for functional enrichment analysis. We expect that after incorporating these genes, functional analysis will provide more significant GO terms and KEGG pathway annotations than the traditional approach of adopting DEGs only. To validate our hypothesis, a knock-down experiment on *birc5a* with a morpholino (Birc5aMO) in zebrafish embryo neuron development and an RNA-seq experiment for detecting DEGs and DE lncRNAs were performed. The detailed procedures and results are introduced and discussed in the following sections.

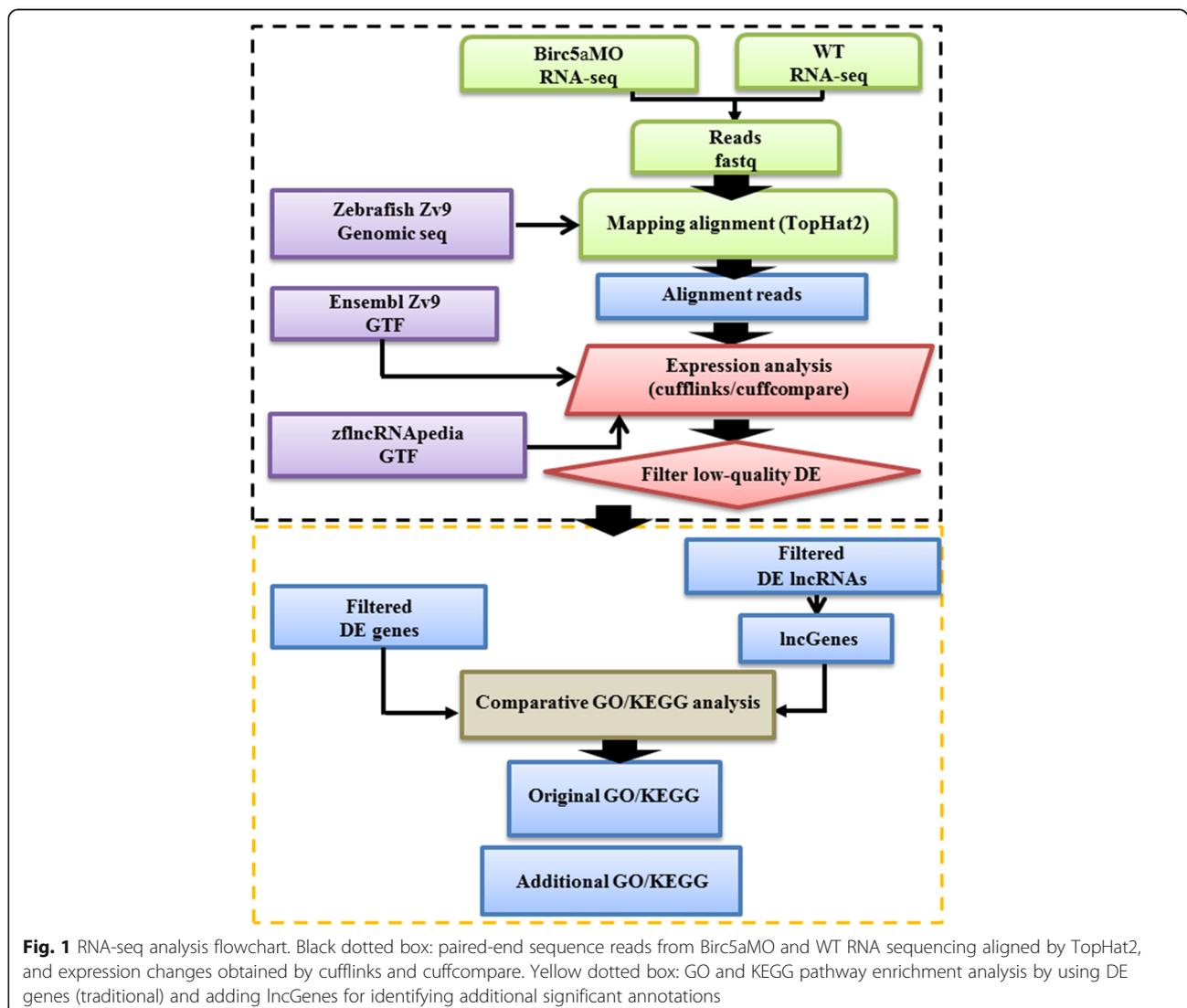
## Methods

### RNA-Seq

The proposed system configuration is shown in Fig. 1. First, *birc5a* morpholino-treated (Birc5aMO) and untreated wild type (WT) zebrafish embryos were collected 20 h post-fertilization (hpf), and the total RNA was extracted from each sample (with two repeats) using TRIzol reagent. After RNA preparation, the cDNA libraries were constructed by Genomics (NTC, Taiwan). The cDNA libraries were generated and sequenced with a NextSeq 500 sequencer (Illumina, CA, USA) with paired-end short read fragments 100 bp in length. The resulting fastq sequence files were further filtered and trimmed to remove contaminating artifacts and low-quality base calls for the following sequence mapping alignment.

### Short Read Alignment and Expression Calculation.

Filtered reads were aligned to a zebrafish genome sequence (Ensembl release 79 version Zv9) by TopHat2 [18]. The mapped reads from bam files were further analyzed and normalized to corresponding expression patterns and values by Cufflinks [19]. The unit of expression value is FPKM (Fragments Per Kilobase of transcript per Million). Gene expression annotation was analyzed with the Ensembl Zv9 transcript GTF file. For lncRNA analysis, the GTF file from zflncRNpedia [20] was applied for unified annotation of lncRNAs in zebrafish. To compare differential expression (DE) of genes and lncRNAs between the Birc5aMO and WT control datasets, Cuffcompare [19] was applied by using the average FPKM of repeat samples. Fold changes in gene and lncRNA expression of  $> 1.5$  or  $< -1.5$  were considered to indicate



**Fig. 1** RNA-seq analysis flowchart. Black dotted box: paired-end sequence reads from Birc5aMO and WT RNA sequencing aligned by TopHat2, and expression changes obtained by cufflinks and cuffcompare. Yellow dotted box: GO and KEGG pathway enrichment analysis by using DE genes (traditional) and adding IncGenes for identifying additional significant annotations

significantly expressed genes in enrichment analysis. Genes nearby or overlapped with DE lncRNAs (within a default parameter of 5 kb upstream or downstream of protein-coding regions) were identified and considered to be DE lncRNA-associated genes (lncGenes).

**GO and KEGG pathway enrichment analysis.**

The annotated DE genes and lncGenes were further explored for their corresponding biological meaning by GO annotations retrieved from the Gene Ontology Consortium (version 2017.03.01) [21] and mapped against biological pathway information in the KEGG database (version 0.7.2).

For GO and KEGG pathway enrichment analysis, we applied hypergeometric distribution statistical theory to calculate *p*-values for each pathway and each GO term. The formula for traditional over-representation analysis is represented in Eq. (1).

$$P_j(\text{at least } i \text{ genes} | N, n, K) = \sum_{k=i}^{\min(n,K)} \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}} \quad (1)$$

where *N* is the number of all unique genes mapped against the reference dataset, *n* is the number of DE genes, *K* is the number of all genes annotated in a specific GO term or a specific KEGG pathway, and *k* is the number of DE genes mapped to the specific GO term or the specific KEGG pathway. The *p*-value for the *j*<sup>th</sup> GO term or the *j*<sup>th</sup> KEGG pathway is denoted by *P<sub>j</sub>*. The proposed novel GO and KEGG pathway enrichment analysis considers the additionally identified DE lncRNAs and their associated genes (lncGenes). Therefore, the modified formula for over-representation analysis is represented in Eq. (2).

$$P_j(\text{at least } i \text{ genes} | N, n, K, r) = \sum_{k=i}^{\min(n+r,K)} \frac{\binom{K}{k+r} \binom{N-K}{(n+r)-(k+r)}}{\binom{N}{n+r}} \quad (2)$$

where *r* is the number of lncGenes and all other parameters (*N*, *n*, *K*, *k*, and *P<sub>j</sub>*) denote the same values as in Eq. (1).

When *P<sub>j</sub>* is less than or equal to 0.05, the specified GO term or KEGG pathway was considered a significant functional annotation (*p*-value <= 0.05). In the GO structure, the identified significant parent GO terms with significant child terms were removed according to the GO tree structure relationship for the most specific and accurate annotations.

**Results**

After performing read quality check procedures on the RNA-seq datasets, a total of 20,759,640 (Birc5aMO repeat 1), 21,003,970 (Birc5aMO repeat 2), 22,221,792 (WT repeat 1), and 22,456,290 (WT repeat 2) reads were retained as high-quality reads for functional analysis. TopHat2 and Cufflinks were used for sequence mapping and differential expression analysis of the Birc5aMO and WT control datasets; two resulting gene sets, including 638 DE genes and 438 DE lncRNAs, were identified with significant fold changes relative to 29,806 unigenes and 17,488 lncRNAs, respectively, based on the FPKM normalization mechanism (see Table 1).

Among the identified 438 DE lncRNAs, 408 novel lncRNAs were found in this experiment; the remaining 30 lncRNAs had already been annotated within the zflncRNAPedia database. To further analyze the DE lncRNAs and potentially associated genes, we searched for overlapped and nearby genes located within 5 kb of the identified DE lncRNAs; a total of 616 neighboring or overlapped genes, called lncGenes, were retrieved from the zebrafish genome. In biological functional interpretations through GO and KEGG pathway enrichment analysis, several significant GO terms and KEGG pathways were additionally identified using the combination of DE genes and lncGenes (Table 2).

When using DE genes and lncGenes, the proposed method detected an additional 99 GO terms compared to the traditional method using DE genes only, of which only 85 GO terms were remained for functional annotations after removing parent GO terms with significant child terms. Similarly, an additional 21 significant KEGG pathways were identified compared to those found by traditional enrichment analysis. In Fig. 2, we selected neuron-related GO terms, which were significantly changed when the additional lncGenes were included in the GO enrichment analysis. When using only DE genes, only one GO terms related to the neuron (GO:0090299) was identified as significant. Comparatively, when additional lncGenes were involved in GO enrichment analysis, four additional GO terms associated with the neuron (GO:0001841, GO:0021879, GO:0048885, and

**Table 1** Different Expression Statistics of RNA-Seq

	Genes	lncRNAs
Total	29,806	17,488
Filter <sup>a</sup>	638	438
> 1.5 Fold change <sup>b</sup>	320	268
< -1.5 Fold change <sup>b</sup>	318	170

<sup>a</sup>Removed FPKM values SD <=15%, fold changes lower than 1.5, or FPKM < 1 If fold change was > 1.5, removed Birc5aMO genes with FPKM less than 1; if fold change was < -1.5, removed WT genes with FPKM less than 1

<sup>b</sup>Compared average FPKM of Birc5aMO/WT from filtered messenger RNAs (mRNAs) and lncRNAs

**Table 2** GO and KEGG enrichment analysis

Annotation	GO <sup>a</sup>	KEGG <sup>a</sup>
Significant items detected from DE genes	165	7
Significant items detected from DE genes + lncGenes <sup>b</sup>	264	28
Additionally detected items	99	21

<sup>a</sup>The results were filtered at  $p$ -value  $\leq 0.05$

<sup>b</sup>Genes associated with DE lncRNAs located at upstream or downstream of the transcription region (within 5 kb)

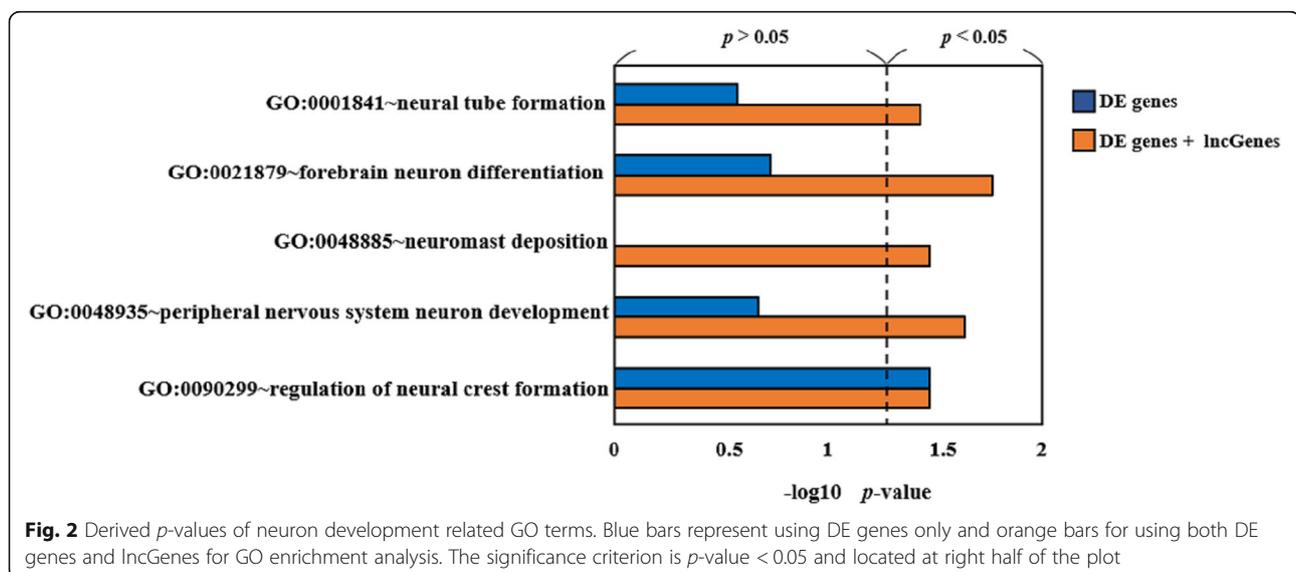
GO:0048935), including neural tube formation, forebrain neuron differentiation, neuromast deposition, and nervous system neuron development, were observed.

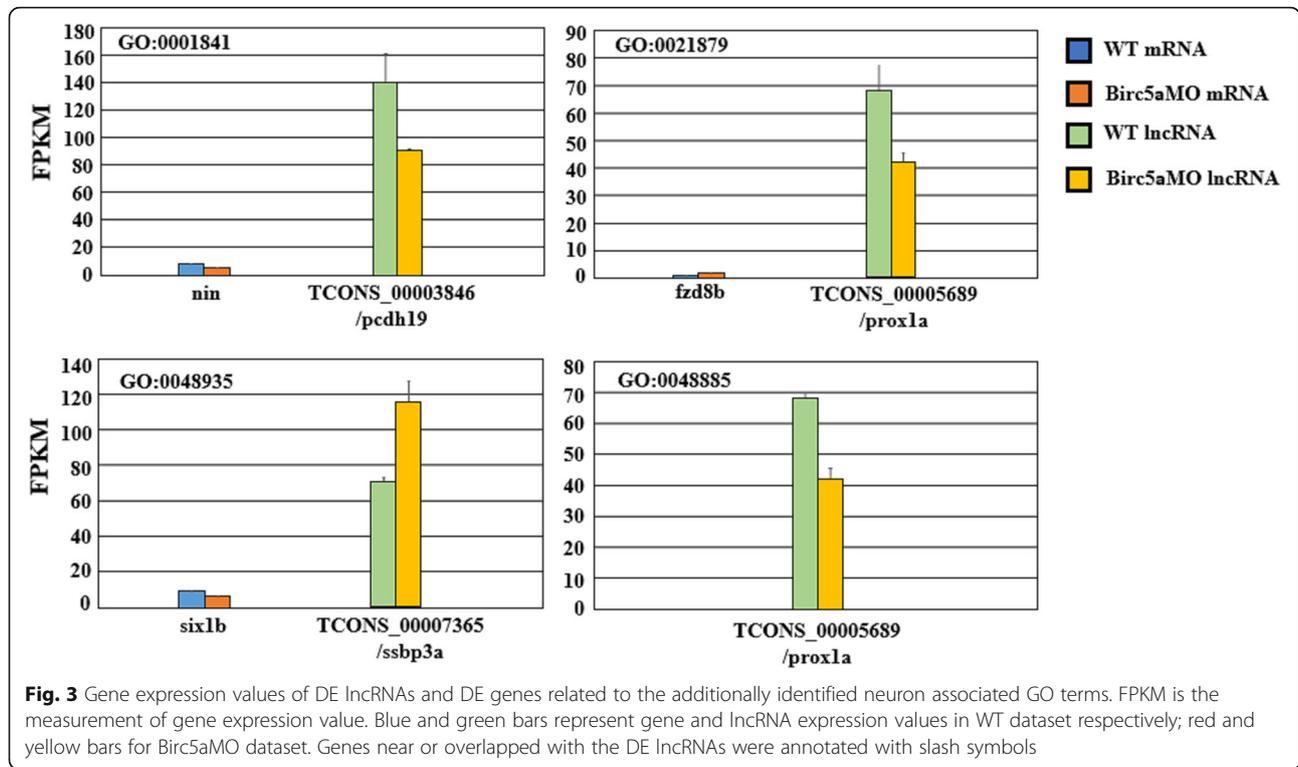
Figure 3 shows expression values of DE lncRNAs and DE genes annotated within the additionally identified GO terms. These four additional GO terms related to neurons and the brain were retrieved by combing DE genes and lncGenes during enrichment analysis. However, the GO:0048885 term only possessed annotated lncGenes; in other words, this GO term contain no DE genes. All annotated DE genes and DE lncRNAs are shown according to their FPKM expression values for the other four additionally identified GO terms. Notably, FPKM expression values of the DE lncRNAs are relatively greater than those of the genes. These phenomena imply that the identified DE lncRNAs act vigorously and that their associated genes may play important regulatory roles during neuron development. Additionally, all illustrated lncRNAs were novel noncoding RNAs (named with TCONS initials). We also observed that the expression levels of the associated lncGenes did not undergo significant changes in expression. Upon further examination of the distances between the DE lncRNAs and their associated lncGenes, one overlapped with a 5'

exon1 segment (TCONS\_00003846), and two overlapped with 3'UTR regions (TCONS\_00007365 and TCONS\_00005689). Additionally, the lncGene *prox1a* was involved with two additionally detected neuron-related GO terms, which indicates that its DE lncRNA might associate with *prox1a* and play an important role in neuron development through the regulation of *birc5a*.

When using DE genes and lncGenes for KEGG pathway enrichment analysis, a number of additional pathways were found compared to those found using DE genes only. The apoptosis-associated pathways in the survivin gene knock-down experiment are illustrated for comparison along with corresponding  $p$ -values in Fig. 4. In investigating signal transduction pathways, only the Notch signaling pathway appeared when only DE genes were used in enrichment analysis. Another six signal transduction pathways, including the important MAPK and mTOR pathways, were retrieved by combining DE genes and lncGenes. When integrating the signal transduction and apoptosis pathways, we observed that the FoxO and MAPK signaling pathways were both major sub-networks directly connected to apoptosis (Fig. 5).

Other signal transduction pathways were dispersed around the MAPK pathway, and were not directly connected to the apoptosis cluster. Twelve DE lncRNAs and associated lncGenes were detected among the apoptosis pathway. In the FoxO and MAPK signaling pathways, only two overlapped lncRNAs, *cacna1sboa1* and *lnc2\_prkra*, are defined in the zflncRNpedia database. The other novel identified DE lncRNAs were categorized as three lincRNAs (TCONS\_00006696, TCONS\_00009550, and TCONS\_00011673), three 3'UTR overlapped lncRNAs (TCONS\_00004392, TCONS\_00009432, and TCONS\_00012098), one 3' exon overlapped lncRNA (TCONS\_00013138), and three 5' UTR and/or exon



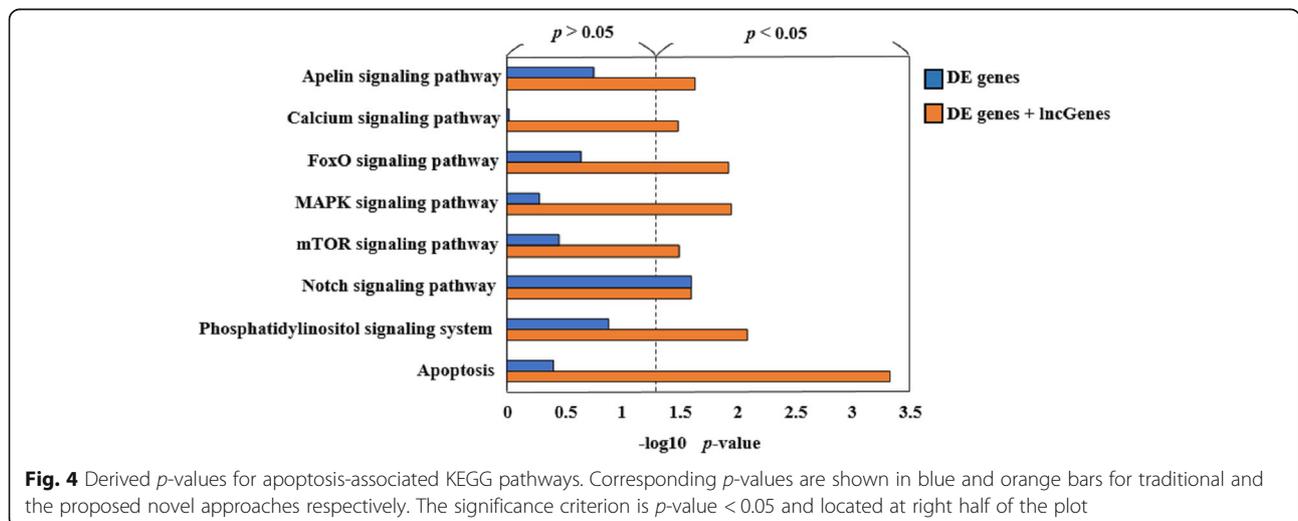


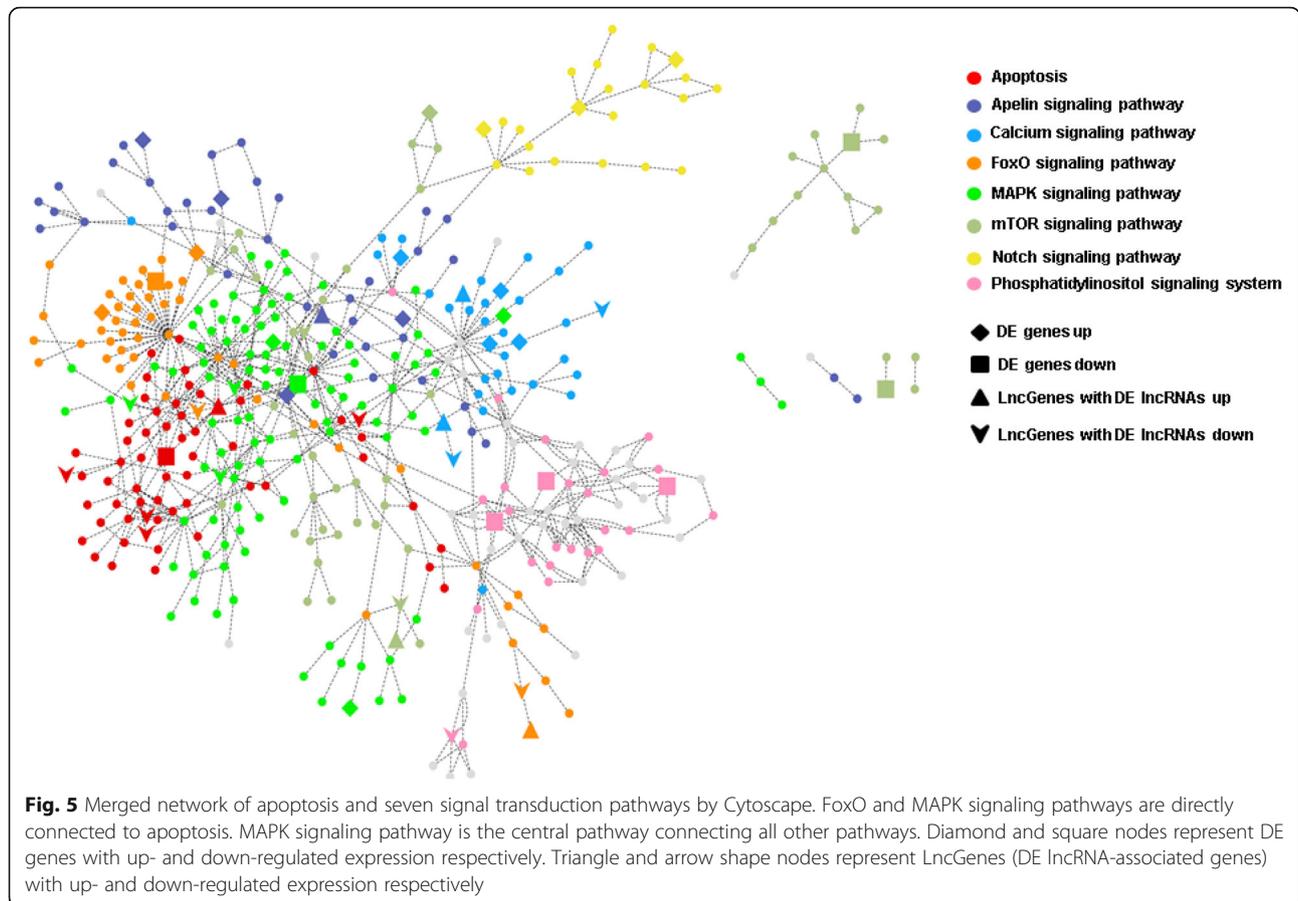
overlapped (TCONS\_00008269, TCONS\_00010719, and TCONS\_00013347). DE genes and lncRNAs associated with DE lncRNAs that were simultaneously involved in significant GO terms and KEGG pathways are listed in Table 3. Three DE genes and three lncRNAs were detected in the survivin knock-down experiments. Among these identified significant functional terms for both KEGG and GO terms, only the Notch signaling pathway and GO:0007219 (Notch signaling pathway) were identified by traditional approach. All

other associated significant pathways and GO terms were identified through our proposed method.

### Discussion

In functional analysis of differential expression gene profiles, a group of DE genes clustered and annotated by an identical GO term or a pathway elucidates the responses and effects of transcription changes based on a hypergeometric distribution. Although this approach completely indicates the associations between biological functions and





affiliated genes, two major disadvantages exist. First, the significant annotations depend on *p*-value calculations, and if the collected functional annotations are not sufficient, no significant annotations can be identified. Second, most genes affected by ncRNAs are not considered; thus, some pertinent annotations could be ignored due to a statistically insignificant *p*-value. Hence, in traditional DE gene analysis, several hidden gene nodes related to specific biological functions could be neglected. To retrieve the hidden nodes within GO terms and KEGG pathways, ncRNA-associated genes have been proposed to enhance

traditional DE gene functional analysis. Our strategy was to consider all transcripts, including mRNAs and ncRNAs, for biological functional annotations. Previous studies reported that lncRNAs were classified into various types: intronic lncRNA, cds overlapping lncRNAs, antisense lncRNAs, and intergenic lncRNAs, among others. [21]. However, no single database provides comprehensive annotations for the various types of lncRNAs on regulatory relationships; notably, the lncRNA target gene database (lncRNAs-RNAs) is highly deficient. The adjacent lncRNAs effecting nearby gene expression in fission yeast, *Schizosaccharomyces*

**Table 3** Overlapped DE genes and lncRNAs in KEGG pathways and GO terms

DE gene	FPKM(WT/Birc5aMO)	KEGG <sup>a</sup>	GO <sup>b</sup>
<i>cacnb4b</i>	6.88/3.7	MAPK signaling	GO:0070588~calcium ion transmembrane transport
<i>Dla</i>	38.61/20.51	Notch signaling	GO:0007219~Notch signaling pathway
<i>impa1</i>	8.02/12.31	Phosphatidylinositol signaling system	GO:0046854~phosphatidylinositol phosphorylation
DE lncRNA / lncGene	FPKM(WT/Birc5aMO)	KEGG	GO
<i>cacna1sboa1 / cacna1sb</i>	5.43/2.69	MAPK signaling	GO:0070588~calcium ion transmembrane transport
TCONS_00007367 / <i>impad1</i>	99.4/163.82	Phosphatidylinositol signaling system	GO:0046854~phosphatidylinositol phosphorylation
TCONS_00008341 / <i>tbc1d7</i>	138.67/208.94	mTOR signaling	GO:0032007~negative regulation of TOR signaling

<sup>a</sup>All pathways are additionally identified except Notch signaling pathway

<sup>b</sup>All GO terms are additionally identified except GO:0007219

*pombe*, has been discussed. The lncRNA *nc-tgp1* increases expression of the neighboring gene *tgp1*<sup>+</sup> (glycerophosphodiester 1), which is involved in glycerophospholipid metabolism [22]. In female mouse embryonic stem cells, transcription and splicing of the lncRNA *Blustr* activates *Sfmbt2* gene expression and affects protein and histone binding [23]. In their research on E11 mouse embryos, Carlson et al. observed that lncRNA-HIT could enhance developmental transcription factors *Hoxa13* and *Hoxa11* [24]. Sarangdhar et al. observed a lncRNA called *durga*, which regulates *kalrna* expression to control dendritic cell length and density in zebrafish [25]. lncRNA regulation may be based on different mechanisms, such as co-regulation, Pol II inhibition, RNA stability, and miRNA regulation [26, 27]. Expression of lncRNAs produced from gene promoter regions probably destabilizes the RNA Pol II transcription-initiation complex, resulting in inhibition of Pol II activity [26]. lncRNAs also regulate mRNA stability; the 3'UTR region of mRNAs may interact with lncRNAs that possess Alu repeat elements, which would induce STAU1-mediated decay (SMD) activities and lead to mRNA decay. Additionally, some lncRNAs such as *bace1as* create RNA-RNA duplexes by combining with mRNAs, and these structure could eliminate miRNA-mediated mRNA instability [28]. The regulatory relationship between lncRNAs and miRNAs can be divided into four categories [29]. First, some lncRNAs are regarded as precursors for miRNA production. Second, miRNAs can bind to lncRNAs directly instead of binding to mRNA. Third, lncRNAs can compete with miRNAs for binding sites on mRNAs. Finally, lncRNAs can act as “sponges” that bind both mRNAs and miRNAs together. lncRNAs can also regulate gene expression by binding to proteins such as transcription factors and help to activate or repress gene expression [30]. Recent studies have shown lncRNAs such as H3K27me3 or H3K4me3 associating with chromatin methylation [13]. Studies have also indicated that lncRNAs bound to proteins involved in chromatin modification, including by methylation of DNA or lysine 9 of histone H3, which control nucleosome structure and result in gene expression changes [31]. lncRNAs bind to chromatin and form protein complexes, and lncRNA-protein complexes interact with methylated chromatin regions to regulate gene expression [32]. Regulation by neighboring lncRNAs has been experimentally demonstrated [23]. Thus, genes overlapped with neighboring DE lncRNAs or within a short range were assumed to be hidden genes participating in functional activities, and the default distance between DE lncRNAs and target genes (lncGenes) was heuristically set at 5 kb as a default value in this study.

To investigate lncRNA expressions and locations in zebrafish, Heena et al. created the lncRNA dataset *zflncRNpedia* which collected a total of 2267 lncRNAs.

However, the dataset only included one RNA-seq datasets from each of the early embryogenesis stage, late developmental stage, and adult tissues. When embryos are stressed under conditions such as hypoxia, some induced or repressed lncRNAs can be observed. Our results indicated that the low expression of *hif2a* in a normoxic environment can invoke *Birc5a* to protect neural progenitor cells and promote neural differentiation [3]. In this study, we discovered additional neuron-related GO terms using both DE genes and lncGenes (DE lncRNA associated genes) in functional analysis. When using only DE genes to identify significant GO terms, only one neuron-related term was found. We observed that most DE lncRNAs possessed higher FPKM levels than did DE genes among the neuron-related GO terms, which may reflect hidden events regarding the GO functions. When applying DE genes and lncGenes simultaneously, *prox1a* was involved in half of the neuron-related GO terms, which indicated that this gene plays an important role in CNS (central neural system) development. In addition, zebrafish *in situ* studies have shown that *prox1a* is expressed in the midbrain-forebrain boundary, hindbrain, and spinal cord neurons [33, 34].

For pathway annotations, when using DE genes exclusively in KEGG pathways analysis, only the Notch signaling pathway was detected as significant. However, several signal transduction pathways and apoptosis pathways appeared when using both DE genes and lncGenes (DE lncRNAs associated genes). Survivin is known to play an essential role in apoptosis-mediated neurogenesis [35], and our study observed *birc* gene orthologues possessing the same function as survivin. In addition, the FoxO signaling pathway affiliates multiple biological functions, including apoptosis, glucose metabolism, and oxidative stress response. Activation of FOXO3/FKHRL (key genes in FoxO signaling pathway) would induce repression of *birc5/survivin* and apoptosis of neuroblastoma cells [36]. Pathway analysis showed FoxO and MAPK were the main pathways connected to apoptotic processes. Three pathways, including MAP/ERK, JNK/p38, and ERK5, belonged to the subnetworks of the MAPK pathway. The DE genes and lncGenes showed that regulation of the MAP/ERK and JNK/p38 pathways is related to cell proliferation and apoptosis. Previous studies have indicated that the MAP/ERK signaling pathway is necessary for CNS development [37], and that the JNK/p38 signaling pathway plays an essential role in the differentiation of neural stem cells [38].

The other signaling transduction pathways located around the MAPK pathway are not connected to apoptosis directly, but their functions are also crucially related to neuron development. Several Notch signaling pathway studies on neuronal development and neuronal

degeneration have been reported [39]. The phosphatidylinositol signaling pathway is important in intra- and extra-cellular signaling and affects neuron development [40]. TOR signaling has also been reported as a key molecular pathway in nervous system development [41]. The calcium signaling pathway is a neuronal-specific signaling pathway [42], and several DE genes and lncGenes annotated in the MAPK/apoptosis pathways are also annotated in GO functional analysis of the calcium signaling pathway. Many studies have observed the effects of calcium on neurons; Duchen reported that neuronal cell death was mediated by calcium signaling [43]. Turner reported that calcium-activated potassium channels expressed in excitatory and inhibitory cells played a role in CNS development [44]. Calcium influxes into neurons increases ATP production, which stimulates the citric acid cycle in mitochondria, supplying energy for neuronal metabolism [45]. These described pathways were identified by using both DE genes and DE lncRNA-associated genes (lncGenes) in KEGG pathway enrichment analysis.

## Conclusions

We have proposed a method that combines expression changes of genes and lncRNAs for biological functional analysis. Our results suggest that ncRNAs play essential roles in regulating gene expression. The retrieval of additional significant GO terms and KEGG pathways provided important context for understanding the regulation of *birc5a* in embryonic neuron development. Further research verifying the target genes of associated DE lncRNAs will be conducted through biological experiments for clarification.

## Abbreviations

Birc5aMO: Birc5a morpholino-treated; DE lncRNA: Differentially expressed long noncoding RNA; DE: Differential expression; DEG: Differentially expressed gene; FPKM: Fragments Per Kilobase of transcript per Million; GO: Gene ontology; lncGene: lncRNA-associated gene; lncRNA: Long noncoding RNA; miRNA: MicroRNA; mRNA: Messenger RNA; ncRNA: Noncoding RNA; WT: Wild type

## Acknowledgements

We thank Mr. Chun-Cheng Liu for his assistance on biological pathway analysis.

## Competing interests

The authors declare that they have no competing interests.

## Funding

This work was supported by the Center of Excellence for the Oceans of the National Taiwan Ocean University and Ministry of Science and Technology, Taiwan (MOST 104–2627-B-019-003, MOST 104–2321-B-019-008, and MOST 105–2313-B-019-001 to Tun-Wen Pai). Publication costs was funded by MOST 104–2627-B-019-003.

## Availability of data and materials

Not applicable

## Authors' contributions

KSH and TWP conceived the algorithm and wrote the manuscript. CCH designed the algorithms and performed GO enrichment analysis. CHH and

WST performed RNA-seq experiments. WDW and YRC proofread and revised the manuscript. All authors read and approved the final manuscript.

## Consent for publication

Not applicable

## Ethics approval and consent to participate

Not applicable

## About this supplement

This article has been published as part of *BMC Systems Biology* Volume 12 Supplement 4, 2018: Selected papers from the 11th International Conference on Systems Biology (ISB 2017). The full contents of the supplement are available online at <https://bmcsystbiol.biomedcentral.com/articles/supplements/volume-12-supplement-4>.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Author details

<sup>1</sup>Department of Computer Science and Engineering, National Taiwan Ocean University, Keelung, Taiwan. <sup>2</sup>Department of Bioscience and Biotechnology, National Taiwan Ocean University, Keelung, Taiwan. <sup>3</sup>Department of Bioagricultural Science, National Chiayi University, Chiayi, Taiwan. <sup>4</sup>Agricultural Biotechnology Research Center, Academia Sinica, Taipei, Taiwan. <sup>5</sup>Center of Excellence for the Oceans, National Taiwan Ocean University, Keelung, Taiwan.

Published: 24 April 2018

## References

- Webster WS, Abela D. The effect of hypoxia in development. *Birth Defects Res C Embryo Today*. 2007;81(3):215–28.
- Corcoran A, O'Connor JJ. Hypoxia-inducible factor signalling mechanisms in the central nervous system. *Acta Physiol (Oxf)*. 2013;208(4):298–310.
- Ko CY, et al. Integration of CNS survival and differentiation by HIF2alpha. *Cell Death Differ*. 2011;18(11):1757–70.
- Sato A, et al. Blocking COX-2 induces apoptosis and inhibits cell proliferation via the Akt/survivin- and Akt/ID3 pathway in low-grade-glioma. *J Neuro-Oncol*. 2017;132(2):231–8.
- Rubio N, Garcia-Segura LM, Arevalo MA. Survivin prevents apoptosis by binding to caspase-3 in astrocytes infected with the BeAn strain of Theiler's murine encephalomyelitis virus. *J Neuro-Oncol*. 2012;18(5):354–63.
- Qureshi IA, Mattick JS, Mehler MF. Long non-coding RNAs in nervous system function and disease. *Brain Res*. 2010;1338:20–35.
- Hart RP, Goff LA. Long noncoding RNAs: central to nervous system development. *Int J Dev Neurosci*. 2016;55:109–16.
- Chou CH, et al. miRBase 2016: updates to the experimentally validated miRNA-target interactions database. *Nucleic Acids Res*. 2016;44(D1):D239–47.
- Vance KW, Ponting CP. Transcriptional regulatory functions of nuclear long noncoding RNAs. *Trends Genet*. 2014;30(8):348–55.
- Ma L, Bajic VB, Zhang Z. On the classification of long non-coding RNAs. *RNA Biol*. 2013;10(6):925–33.
- Anderson DM, et al. A micropeptide encoded by a putative long noncoding RNA regulates muscle performance. *Cell*. 2015;160(4):595–606.
- Nelson BR, et al. A peptide encoded by a transcript annotated as long noncoding RNA enhances SERCA activity in muscle. *Science*. 2016; 351(6270):271–5.
- Pauli A, et al. Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. *Genome Res*. 2012;22(3):577–91.
- Liao Q, et al. ncFANS: a web server for functional annotation of long non-coding RNAs. *Nucleic Acids Res*. 2011;39(Web Server issue):W118–24.
- Park C, et al. lncRNator: a comprehensive resource for functional investigation of long non-coding RNAs. *Bioinformatics*. 2014;30(17):2480–5.
- Zhao Q, et al. Co-LncRNA: investigating the lncRNA combinatorial effects in GO annotations and KEGG pathways based on human RNA-Seq data. *Database (Oxford)*. 2015;2015.
- Serin EA, et al. Learning from co-expression networks: possibilities and challenges. *Front Plant Sci*. 2016;7:444.

18. Kim D, et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 2013;14(4):R36.
19. Trapnell C, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and cufflinks. *Nat Protoc.* 2012;7(3):562–78.
20. Dhiman H, et al. zlncRNpedia: a comprehensive online resource for zebrafish long non-coding RNAs. *PLoS One.* 2015;10(6):e0129997.
21. Chakraborty S, et al. LncRBase: an enriched resource for lncRNA information. *PLoS One.* 2014;9(9):e108010.
22. Ard R, Allshire RC. Transcription-coupled changes to chromatin underpin gene silencing by transcriptional interference. *Nucleic Acids Res.* 2016;44(22):10619–30.
23. Engreitz JM, et al. Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature.* 2016;539(7629):452–5.
24. Carlson HL, et al. lncRNA-HIT functions as an epigenetic regulator of Chondrogenesis through its recruitment of p100/CBP complexes. *PLoS Genet.* 2015;11(12):e1005680.
25. Sarangdhar MA, et al. A novel long non-coding RNA, durga modulates dendrite density and expression of kalirin in zebrafish. *Front Mol Neurosci.* 2017;10:95.
26. Kung JT, Colognori D, Lee JT. Long noncoding RNAs: past, present, and future. *Genetics.* 2013;193(3):651–69.
27. Mercer TR, et al. Expression of distinct RNAs from 3' untranslated regions. *Nucleic Acids Res.* 2011;39(6):2393–403.
28. Geisler S, Collier J. RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts. *Nat Rev Mol Cell Biol.* 2013;14(11):699–712.
29. Barry G. Integrating the roles of long and small non-coding RNA in brain function and disease. *Mol Psychiatry.* 2014;19(4):410–6.
30. Bassett AR, et al. Considerations when investigating lncRNA function in vivo. *elife.* 2014;3:e03058.
31. Bohmdorfer G, Wierzbicki AT. Control of chromatin structure by long noncoding RNA. *Trends Cell Biol.* 2015;25(10):623–32.
32. Fatica A, Bozzoni I. Long non-coding RNAs: new players in cell differentiation and development. *Nat Rev Genet.* 2014;15(1):7–21.
33. Sprague J, et al. The zebrafish information network: the zebrafish model organism database provides expanded support for genotypes and phenotypes. *Nucleic Acids Res.* 2008;36(Database issue):D768–72.
34. Glasgow E, Tomarev SI. Restricted expression of the homeobox gene *prox 1* in developing zebrafish. *Mech Dev.* 1998;76(1–2):175–8.
35. Jiang Y, et al. Essential role for survivin in early brain development. *J Neurosci.* 2005;25(30):6962–70.
36. Obexer P, et al. Repression of BIRC5/survivin by FOXO3/FKHRL1 sensitizes human neuroblastoma cells to DNA damage-induced apoptosis. *Mol Biol Cell.* 2009;20(7):2041–8.
37. Samuels IS, Saitta SC, Landreth GE. MAP'ing CNS development and cognition: an ERKsome process. *Neuron.* 2009;61(2):160–7.
38. Yang SR, et al. The role of p38 MAP kinase and c-Jun N-terminal protein kinase signaling in the differentiation and apoptosis of immortalized neural stem cells. *Mutat Res.* 2005;579(1–2):47–57.
39. Pope GR, et al. Central and peripheral apelin receptor distribution in the mouse: species differences with rat. *Peptides.* 2012;33(1):139–48.
40. Lo Vasco VR. Phosphoinositide pathway and the signal transduction network in neural development. *Neurosci Bull.* 2012;28(6):789–800.
41. Tee AR, et al. The role of mTOR signalling in neurogenesis, insights from tuberous sclerosis complex. *Semin Cell Dev Biol.* 2016;52:12–20.
42. Khodosevich K, Monyer H. Signaling in migrating neurons: from molecules to networks. *Front Neurosci.* 2011;5:28.
43. Duchon MR. Mitochondria, calcium-dependent neuronal death and neurodegenerative disease. *Pflugers Arch.* 2012;464(1):111–21.
44. Turner RW, et al. Neuronal expression of the intermediate conductance calcium-activated potassium channel *KCa3.1* in the mammalian central nervous system. *Pflugers Arch.* 2015;467(2):311–28.
45. Brini M, et al. Neuronal calcium signaling: function and dysfunction. *Cell Mol Life Sci.* 2014;71(15):2787–814.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

