Identification of novel long non-coding RNA in diffuse intrinsic pontine gliomas by expression profile analysis

YUEHUI LIU[1], HAIPING LIU[2] and DONGWEI ZHANG[1]

[1]Department of Neurology, The Affiliated Hospital of Inner Mongolia University for The Nationalities, Tongliao, Inner Mongolia 028007; [2]College of Life Science, Inner Mongolia University for The Nationalities, Tongliao, Inner Mongolia 028000, P.R. China

Abstract. Diffuse intrinsic pontine glioma (DIPG) is one of the most devastating types of pediatric cancer. Accumulating evidence suggests that the dysregulated expression of long non-coding (lnc)-RNAs is associated with various pathologies of the CNS. However, the expression patterns and prognostic roles of lncRNAs in DIPG have not yet been systematically determined. In the present study, lncRNA expression profiles were obtained from the Gene Expression Omnibus (GEO) database using the lncRNA-mining approach and a differential expression analysis for lncRNAs was performed between DIPG and low-grade brainstem glioma and DIPG and normal pediatric brainstem tissue. Using a two-tailed t-test, 58 and 197 lncRNAs were found to be significantly deferentially expressed (Fold change >2 or <0.5, FDR adjusted P<0.05). To identify the prognostic value of these 255 differentially expressed lncRNAs, univariate and multivariate Cox proportional hazards regression analysis were performed and a 9-lncRNA signature as a potential biomarker for predicting the prognosis of DIPG was constructed. Kaplan-Meier curve analysis showed that patients in the high-risk group exhibited a reduced survival time compared with patients in the low-risk group (median survival of 230 vs. 460 days, log-rank test P<0.001). Moreover, this lncRNA-signature could be used as an independent prognostic marker for DIPG patient survival. The present study provided novel candidates for the investigation of potential diagnostic or prognostic biomarkers and/or therapeutic targets of DIPG, as well as a novel insight into the underlying mechanisms of DIPG.

Correspondence to: Professor Dongwei Zhang, Department of Neurology, The Affiliated Hospital of Inner Mongolia University for The Nationalities, 1742 Holin he Street, Tongliao, Inner Mongolia 028007, P.R. China
E-mail: dweizh@sina.com

Introduction

Diffuse intrinsic pontine glioma (DIPG) is one of the most devastating pediatric cancers, and accounts for 10-15 of pediatric brain and central nervous system (CNS) tumors (1,2). The standard treatment for DIPG currently includes neurosurgery, radiotherapy and chemotherapy. However, the prognosis for DIPG remains poor, due to high relapse rates and rapid progression (3). The 1-, 2- and 5-year survival rates of patients with DIPG are approximately 30%, <10 and <1%, respectively (2,4). Therefore, there is an urgent need to identify novel DIPG-related molecular factors and therapeutic targets for the treatment of DIPG.

It is now commonly accepted that at least 90% of the human genome is actively transcribed, whereas only <2% encodes proteins; the majority of the genome can be transcribed into non-coding RNAs (ncRNAs) (5). ncRNAs can be classified into two major classes based on transcript size: Small ncRNAs (such as microRNAs) and long non-coding RNAs (lncRNAs). To date, thousands of lncRNAs have been identified in humans and other species (6). lncRNAs are commonly defined as RNA molecules longer than 200 nucleotides that are not necessarily translated into proteins (7). Accumulating evidence suggests that lncRNAs play important roles in various biological processes by negatively or positively regulating gene expression at the epigenetic, transcriptional and post-transcriptional levels (8-11). With advances in transcriptome profiling, aberrant lncRNA expression has been observed in various human diseases, including cancer. These dysregulated lncRNAs have been implicated in cancer pathogenesis and development (12-17). Recently, the regulatory roles of lncRNAs have been demonstrated in the nervous system function, and their dysregulated expression is involved in various pathologies of the CNS (18,19). However, the expression patterns and prognostic roles of lncRNAs in DIPG have not yet been systematically determined.

This study aimed to identify lncRNA expression patterns in DIPG compared with brainstem low-grade glioma and normal pediatric brainstem tissue, and identify the lncRNAs associated with the survival of patients with DIPG.

Materials and methods

Datasets. The human microarray dataset GSE26576 (1) was downloaded from the NCBI Gene Expression Omnibus

(GEO) database (www.ncbi.nlm.nih.gov/geo/). The GSE26576 microarray dataset was generated with the GPL570 platform (Affymetrix Human Genome U133 Plus 2.0 Array) and included 26 DIPG samples, 6 brainstems low-grade glioma samples and 2 normal pediatric brainstem samples.

lncRNA expression profiles. lncRNA expression profiles included in the GSE26576 dataset were obtained by repurposing microarray probes using GATExplorer software, as previously described (20,21). Briefly, a series of R packages in GATExplorer software were used to map the data and annotate the lncRNA microarray probes. lncRNA probes that mapped to the human and mouse genomes (derived from the RNAdb database) (22) were retained. Finally, 5635 lncRNAs were identified for further analysis.

Preprocessing and analysis of expression profiles. The raw microarray dataset (CEL file) was obtained from the GEO database and normalized using the Robust Multichip Average (RMA) method, which involved three main steps: Background correction, quantile normalization and log2-transformation. For determination of lncRNA differential expression profiles, a two-tailed T-test was used to identify differentially expressed lncRNAs between patients with DIPG and normal controls, and between patients with DIPG and low-grade glioma. lncRNAs with an adjusted $P<0.05$ after FDR correction and a fold change of >2 or <0.5 were considered as differentially expressed lncRNAs. Hierarchical clustering analysis was performed for the expression data of the differentially expressed lncRNAs using the R package 'pheatmap'.

Statistical analysis. The association between the lncRNA gene expression and patient survival was assessed by univariate Cox regression analysis. The Kaplan-Meier method and two-sided log-rank test were used to compare survival differences between low- and high-risk groups. Multivariate Cox analysis was used to test whether the lncRNA expression signature was independent of other clinical features. Time-dependent receiver operating characteristic (ROC) curves were used to compare the sensitivity and specificity of the lncRNA expression signature for survival prediction.

Functional enrichment analysis. Expression correlation between protein-coding genes and lncRNAs was measured using Pearson correlation coefficients. Functional enrichment analysis was conducted for the protein-coding genes co-expressed with the lncRNAs in GO and KEGG using the ClueGO plugin (version 2.3.3) in Cytoscape (23), and DAVID (david.ncifcrf.gov/, version 6.8) (24). GO terms and KEGG pathways were considered significantly enriched when $P<0.05$.

Results

Identification of differentially expressed lncRNAs between patients with DIPG and normal controls. To identify differentially expressed lncRNAs between patients with DIPG and normal controls, we performed differential expression analysis for lncRNAs using student's t-test. A total of 58 lncRNAs were identified as differentially expressed between patients with DIPG and normal controls

Table I. Nine long non-coding RNAs significantly associated with the survival in the diffuse intrinsic pontine glioma dataset.

Gene symbol	Coefficient	Hazard ratio	Z-score	P-value
AF086217	0.969	2.635	2.517	0.012
AF086391	1.444	4.236	2.439	0.015
AF119852	-1.959	0.141	-2.414	0.016
AK021535	0.595	1.813	2.650	0.008
AK022370	-0.684	0.505	-2.011	0.044
AL050068	-0.947	0.388	-2.042	0.041
BC012548	0.428	1.534	2.061	0.039
BC041658	0.634	1.885	1.983	0.047
AF086127	-0.769	0.464	-2.247	0.025

(Fold change >2 or <0.5, $P<0.05$ after FDR adjustment). Of these, 41 lncRNAs were upregulated, and 17 downregulated, in patients with DIPG.

Identification of differentially expressed lncRNAs between patients with DIPG and low-grade glioma. We performed a differential expression analysis of lncRNA expression profiles between patients with DIPG and low-grade glioma, and identified 197 differentially expressed lncRNAs using student's t-test. (Fold change >2 or <0.5, $P<0.05$ after FDR adjustment). Among the differentially expressed lncRNAs, 125 were upregulated and 72 were downregulated in patients with DIPG.

To demonstrate the significance of the dysregulated lncRNAs in discriminating between patients with DIPG, normal controls and patients with low-grade glioma, we performed an unsupervised hierarchical clustering analysis for all samples according to the expression values of the identified differentially expressed lncRNAs. As shown in Fig. 1, three distinct sample clusters were obtained by hierarchical clustering analysis, suggesting that the 255 differentially expressed lncRNAs were closely associated with DIPG, and could be used to distinguish DIPG from normal brainstem tissue and low-grade glioma.

Identification of an lncRNA expression signature for survival prediction in patients with DIPG. To identify survival-related lncRNAs, we performed univariate Cox proportional hazards regression analysis for the aforementioned 255 differentially expressed lncRNAs. A total of 14 lncRNAs were significantly associated with DIPG patient survival. We conducted a multivariate Cox regression analysis for the 14 survival-related lncRNAs, and identified a set of 9 lncRNAs that were independently associated with the DIPG patient survival time (Table I). We constructed a lncRNA expression signature as a classifier for survival prediction according to the expression of the 9 lncRNAs weighted by the multivariate Cox regression coefficient, as follows: Risk Score=(-3.92)*AF086127 + 1.52*AF086217 + 2.42*AF086391 + (-4.06)*AF119852 + 0.80*AK021535 + (-0.82)*AK022370 + 1.38*AL050068 + (-0.89)*BC012548 + (-2.39)*BC041658. The risk score for each patient was calculated based on the lncRNA gene expression signature. Using the
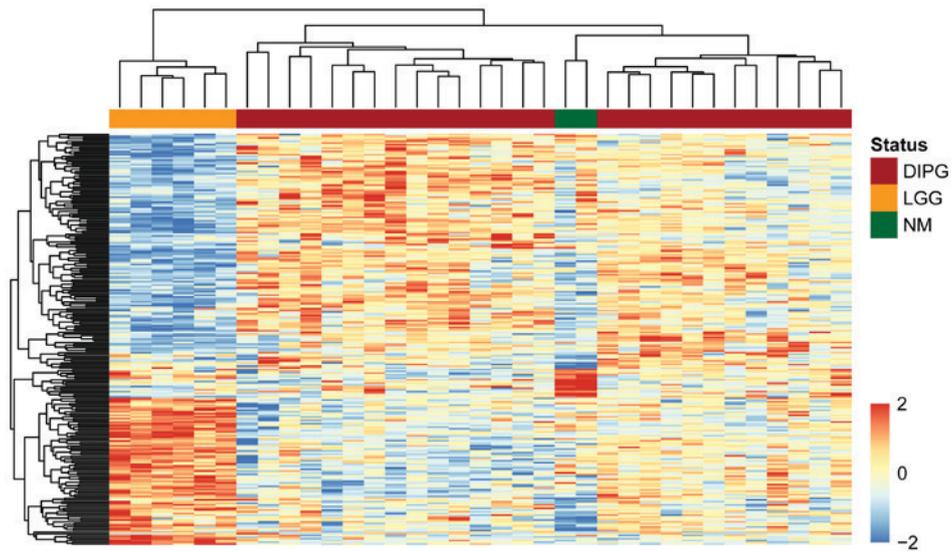
Figure 1. Hierarchical clustering analysis of DIPG, normal control and LGG samples based on 255 differentially expressed lncRNAs. DIPG, diffuse intrinsic pontine glioma; LGG, low-grade glioma; NM, normal brainstem tissue.
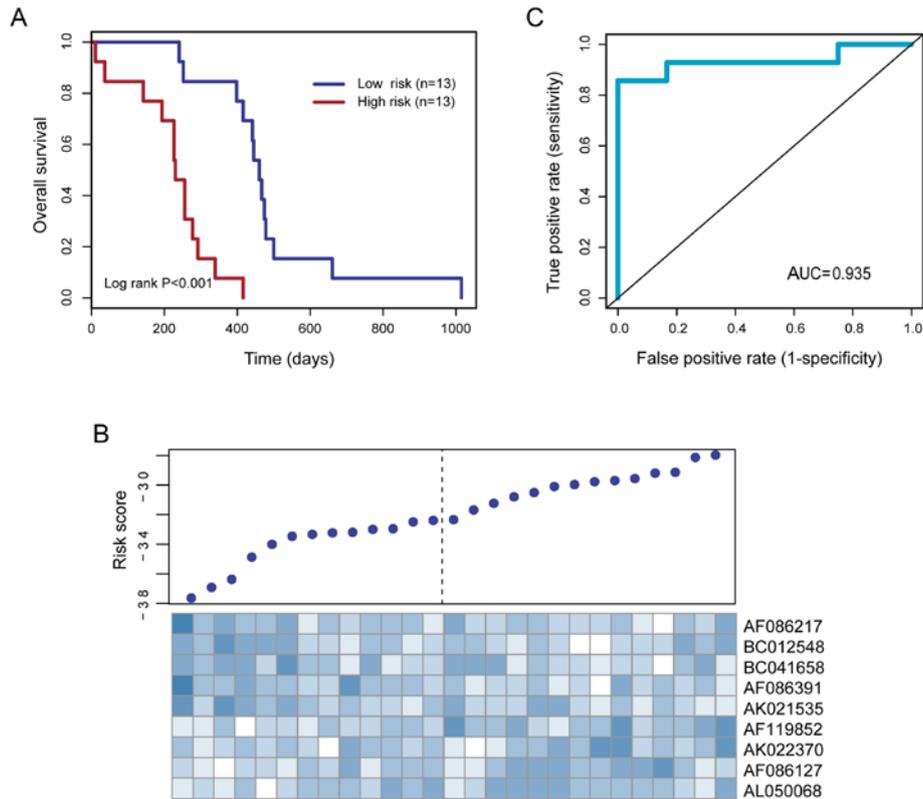

Figure 2. Prognostic performance of the nine-lncRNA signature. (A) Kaplan-Meier curves for the high-risk and low-risk groups. (B) Time-dependent receiver operating characteristic curve analysis of the lncRNA expression signature. (C) Distribution of the risk score and expression heatmap of the nine-lncRNA signature. lncRNA, long non-coding RNA.

median risk score as the cutoff point (-32.36), 26 DIPG patients were classified into the high- and low-risk groups. Patients in the high-risk group exhibited a poorer overall survival time than patients in the low-risk group (median survival of 230 vs. 460 days, log-rank test P<0.001). Kaplan-Meier curves for the high- and low-risk groups are shown in Fig. 2A. The heatmap shows that five protective lncRNAs exhibit a high expression level in the low-risk group, while four risk lncRNAs

exhibit a high expression level in the high-risk group (Fig. 2B). Analysis of time-dependent ROC demonstrated that the AUC value for the lncRNA expression signature was 0.935 for 12-month survival (Fig. 2C).

The 1-year survival rate in the high-risk group was 7.69%, whereas the corresponding rate in the low-risk group was 84.62%. The results of the univariate analysis indicated that the hazard ratio of the high-risk score vs. the low-risk score

Table II. Univariate and multivariate Cox regression analysis of survival in the DIPG dataset.

Variable [DIPG dataset (n=27)]	Univariate analysis			Multivariate analysis		
	HR	95% CI of HR	P-value	HR	95% CI of HR	P-value
Risk score	2.72	1.76-4.21	7.15×10^{-6}	2.69	1.74-4.18	1.01×10^{-5}
Age	1.04	0.93-1.16	0.53	1.01	0.90-1.14	0.84

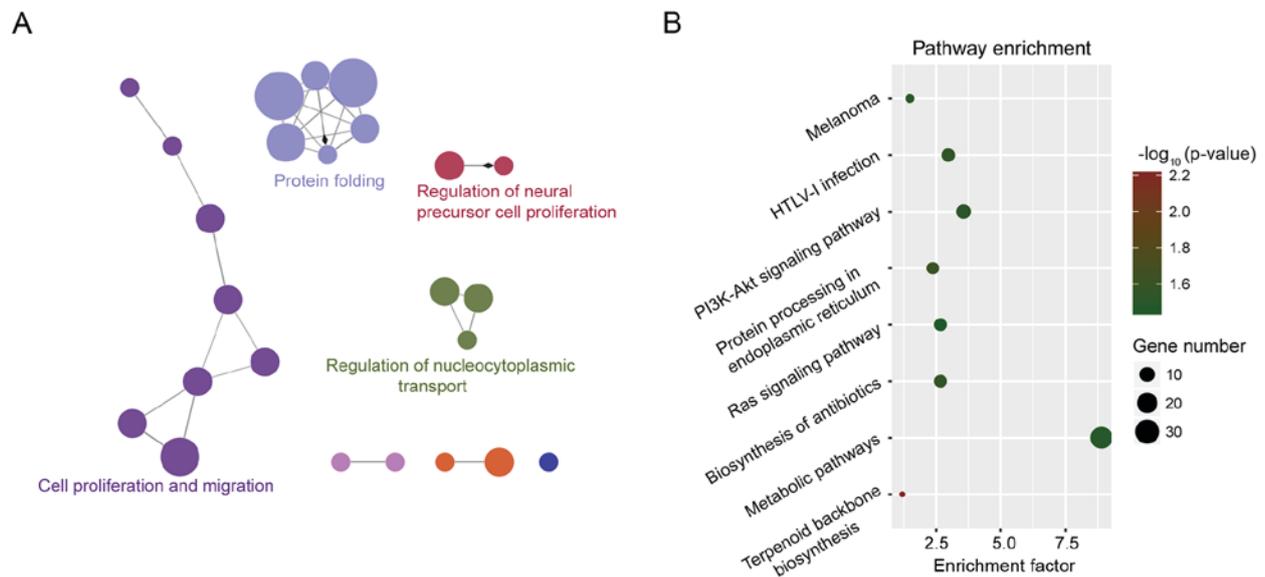HR, hazard ratio; CI, confidence interval; DIPG, diffuse intrinsic pontine glioma.

Figure 3. Functional enrichment analysis. (A) Enriched Gene Ontology terms. (B) Enriched Kyoto Encyclopedia of Genes and Genomes pathways.

for survival was 2.72 [P<0.001; 95% confidence interval (CI)=1.76-4.21] (Table II). According to the multivariate analysis, including age, the hazard ratio for the high-risk vs. the low-risk score for survival was 2.69 (P<0.001; 95% CI, 1.74-4.18) (Table II), indicating that the lncRNA expression signature maintained an independent association with survival.

Functional analysis of the lncRNA expression signature. We performed GO and KEGG enrichment analyses for the protein-coding genes which were co-expressed with the 9 lncRNAs in the gene expression signature using DAVID and clueGO. The results of GO enrichment analysis revealed four enriched GO functional clusters, including 'protein folding', 'cell proliferation', 'epithelial cell migration' and 'regulation of nucleocytoplasmic transport' (Fig. 3A). The results of the KEGG enrichment analysis revealed eight enriched KEGG pathways, including 'terpenoid backbone biosynthesis', 'protein processing in the endoplasmic reticulum', 'biosynthesis of antibiotics', 'HTLV-I infection', 'PI3K-Akt signaling pathway', 'melanoma', 'metabolic pathways' and 'Ras signaling pathway' (Fig. 3B).

Discussion

DIPGs, representing 75-80% of pediatric brainstem tumors, are the most common brainstem tumors in children (25). Previous studies have investigated the molecular heterogeneity between DIPGs and adult high-grade gliomas (HGGs) and between DIPGs and low-grade brainstem gliomas, to improve our understanding of the molecular mechanisms and molecular expression signatures underlying DIPG. Using polymerase chain reaction-single strand polymorphism and nucleotide analyses, Zhang *et al* (26) reported a p53 gene mutation in some DIPGs and inferred that DIPGs might be associated with mutagenic or carcinogenic agents. Paugh *et al* (1) performed genome-wide analyses and demonstrated significantly different frequencies of specific large-scale and local imbalances in gene expression between DIPGs and nonbrainstem pediatric glioblastomas. Another study performed by Lulla *et al* (27) studied the miRNA expression pattern in TPG, and identified two distinct subgroups with differentially expressed microRNAs. A recent study indicated that H3K27M-mutant gliomas share similar histological features and an adverse prognosis in adults and children (28). However, the above studies have focused on genomic mutations, mRNAs or miRNAs. Recent studies have suggested that lncRNAs, a new class of ncRNAs, is an important component of disease biology, and the dysregulated expression of lncRNAs has been observed in various human diseases (29). However, the expression patterns of lncRNAs and their functional roles in DIPGs have not been systematically studied yet.

In this study, we first obtained lncRNA expression profiles of DIPG, brainstem low-grade glioma and normal pediatric brainstem using the lncRNA-mining approach. We

subsequently performed differential expression analysis and identified 58 and 197 significantly differentially expressed lncRNAs between patients with DIPG and normal controls, and between patients with DIPG and low-grade glioma, respectively. To the best of our knowledge, our study is the first to attempt to identify the dysregulated lncRNA expression pattern in patients with DIPG compared with normal controls and patients with low-grade glioma. We hypothesize that these differentially expressed lncRNAs in patients with DIPG may be involved in the pathogenesis and development of DIPG, and could be used as candidates for the investigation of potential diagnostic or prognostic biomarkers and/or therapeutic targets for the treatment of DIPG. During the initial phase of marker discovery, we performed univariate and multivariate Cox proportional hazards regression analysis for these 255 differentially expressed lncRNAs, and constructed a 9-lncRNA signature as a potential biomarker for prognosis of DIPG. Kaplan-Meier curve analysis also demonstrated that patients with the high-risk lncRNA signature had much poorer survival than those with the low-risk lncRNA signature. As radiotherapy and chemotherapy affect the prognosis of patients with DIPG, whether the 9-lncRNA signature is affected by radiotherapy and chemotherapy needs to be investigated in future studies.

Although a large number of lncRNAs have been discovered in humans and animals, few lncRNAs have been functionally characterized. It has been reported that it is an effective method to infer the function of lncRNAs based on coding genes that are co-expressed with these lncRNAs (30). Based on this assumption, we first identified coding genes that are co-expressed with lncRNAs using Pearson correlation coefficients. Functional enrichment analysis was conducted for the protein-coding genes co-expressed with lncRNAs to predict the functions of the 9-lncRNA signature. Functional analysis suggested that the 9-lncRNA signature may be involved in known cancer-related biological pathways and processes. For example, altered Ras signaling has been detected in a variety of cancers, including CNS tumors (31). The PI3K-Akt signaling pathway is well known to be involved in various cellular functions, including nutrient uptake, cell proliferation, growth, autophagy, apoptosis and migration (32), and the dysregulation of the PI3K-Akt signaling pathway is associated with neurodevelopmental disorders (33). In conclusion, we have identified some novel differentially expressed lncRNAs in DIPG using previously generated microarray data and identified a lncRNA signature comprising nine lncRNAs (*AF086127, AF086217, AF086391, AF119852, AK021535, AK022370, AL050068, BC012548* and *BC041658*), which can be collectively used as an independent prognostic marker of DIPG patient survival. Our study provided basis for the further investigation of the mechanisms underlying DIPG.

Acknowledgements

Not applicable.

Funding

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Authors' contributions

DZ conceived and designed the experiments, and wrote the paper. YL and HL performed the experiments and analyzed the data. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Patient consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Paugh BS, Broniscer A, Qu C, Miller CP, Zhang J, Tatevossian RG, Olson JM, Geyer JR, Chi SN, da Silva NS, *et al*: Genome-wide analyses identify recurrent amplifications of receptor tyrosine kinases and cell-cycle regulatory genes in diffuse intrinsic pontine glioma. J Clin Oncol 29: 3999-4006, 2011.
2. Korones DN: Treatment of newly diagnosed diffuse brain stem gliomas in children: In search of the holy grail. Expert Rev Anticancer Ther 7: 663-674, 2007.
3. Jansen MH, van Vuurden DG, Vandertop WP and Kaspers GJ: Diffuse intrinsic pontine gliomas: A systematic update on clinical trials and biology. Cancer Treat Rev 38: 27-35, 2012.
4. Bredlau AL and Korones DN: Diffuse intrinsic pontine gliomas: Treatments and controversies. Adv Cancer Res 121: 235-259, 2014.
5. Stein LD: Human genome: End of the beginning. Nature 431: 915-916, 2004.
6. Brosnan CA and Voinnet O: The long and the short of noncoding RNAs. Curr Opin Cell Biol 21: 416-425, 2009.
7. Spizzo R, Almeida MI, Colombatti A and Calin GA: Long non-coding RNAs and cancer: A new frontier of translational research? Oncogene 31: 4577-4587, 2012.
8. Kornienko AE, Guenzl PM, Barlow DP and Pauler FM: Gene regulation by the act of long non-coding RNA transcription. BMC Biol 11: 59, 2013.
9. Cao J: The functional role of long non-coding RNAs and epigenetics. Biol Proced Online 16: 11, 2014.
10. Zhou M, Wang X, Li J, Hao D, Wang Z, Shi H, Han L, Zhou H and Sun J: Prioritizing candidate disease-related long non-coding RNAs by walking on the heterogeneous lncRNA and disease network. Mol Biosyst 11: 760-769, 2015.
11. Sun J, Shi H, Wang Z, Zhang C, Liu L, Wang L, He W, Hao D, Liu S and Zhou M: Inferring novel lncRNA-disease associations based on a random walk model of a lncRNA functional similarity network. Mol Biosyst 10: 2074-2081, 2014.
12. Zhou M, Wang X, Shi H, Cheng L, Wang Z, Zhao H, Yang L and Sun J: Characterization of long non-coding RNA-associated ceRNA network to reveal potential prognostic lncRNA biomarkers in human ovarian cancer. Oncotarget 7: 12598-12611, 2016.
13. Zhou M, Sun Y, Sun Y, Xu W, Zhang Z, Zhao H, Zhong Z and Sun J: Comprehensive analysis of lncRNA expression profiles reveals a novel lncRNA signature to discriminate nonequivalent outcomes in patients with ovarian cancer. Oncotarget 7: 32433-32448, 2016.

14. Zhou M, Zhao H, Wang Z, Cheng L, Yang L, Shi H, Yang H and Sun J: Identification and validation of potential prognostic lncRNA biomarkers for predicting survival in patients with multiple myeloma. J Exp Clin Cancer Res 34: 102, 2015.

15. Zhou M, Zhao H, Xu W, Bao S, Cheng L and Sun J: Discovery and validation of immune-associated long non-coding RNA biomarkers associated with clinically molecular subtype and prognosis in diffuse large B cell lymphoma. Mol Cancer 16: 16, 2017.

16. Tang JY, Lee JC, Chang YT, Hou MF, Huang HW, Liaw CC and Chang HW: Long noncoding RNAs-related diseases, cancers and drugs. ScientificWorldJournal 2013: 943539, 2013.

17. Zhou M, Zhang Z, Zhao H, Bao S and Sun J: A novel lncRNA-focus expression signature for survival prediction in endometrial carcinoma. BMC Cancer 18: 39, 2018.

18. Qureshi IA, Mattick JS and Mehler MF: Long non-coding RNAs in nervous system function and disease. Brain Res 1338: 20-35, 2010.

19. Zhou M, Zhang Z, Zhao H, Bao S, Cheng L and Sun J: An immune-related six-lncRNA signature to improve prognosis prediction of glioblastoma multiforme. Mol Neurobiol 55: 3684-3697, 2017.

20. Fan Y, Wang YF, Su HF, Fang N, Zou C, Li WF and Fei ZH: Decreased expression of the long noncoding RNA LINC00261 indicate poor prognosis in gastric cancer and suppress gastric cancer metastasis by affecting the epithelial-mesenchymal transition. J Hemat Oncol 9: 57, 2016.

21. Hu Y, Chen HY, Yu CY, Xu J, Wang JL, Qian J, Zhang X and Fang JY: A long non-coding RNA signature to improve prognosis prediction of colorectal cancer. Oncotarget 5: 2230-2242, 2014.

22. Pang KC, Stephen S, Engstrom PG, Tajul-Arifin K, Chen W, Wahlestedt C, Lenhard B, Hayashizaki Y and Mattick JS: RNAdb-a comprehensive mammalian noncoding RNA database. Nucleic Acids Res 33 (Database Issue): D125-D130, 2005.

23. Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, Fridman WH, Pagès F, Trajanoski Z and Galon J: ClueGO: A Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics 25: 1091-1093, 2009.

24. Huang da W, Sherman BT and Lempicki RA: Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4: 44-57, 2009.

25. Warren KE: Diffuse intrinsic pontine glioma: Poised for progress. Front Oncol 2: 205, 2012.

26. Zhang S, Feng X, Koga H, Ichikawa T, Abe S and Kumanishi T: p53 gene mutations in pontine gliomas of juvenile onset. Biochem Biophys Res Commun 196: 851-857, 1993.

27. Lulla RR, Laskowski J, Goldman S, Gopalakrishnan V and Fangusaro J: Microrna profiling reveals two subgroups of diffuse intrinsic pontine glioma. Neuro-Oncology 16: i40–i59, 2014.

28. Kleinschmidt-DeMasters BK and Mulcahy Levy JM: H3 K27M-mutant gliomas in adults vs. children share similar histological features and adverse prognosis. Clin Neuropathol 37: 53-63, 2018.

29. Harries LW: Long non-coding RNAs and human disease. Biochem Soc Trans 40: 902-906, 2012.

30. Ma H, Hao Y, Dong X, Gong Q, Chen J, Zhang J and Tian W: Molecular mechanisms and function prediction of long noncoding RNA. ScientificWorldJournal 2012: 541786, 2012.

31. Fernandez-Medarde A and Santos E: Ras in cancer and developmental diseases. Genes Cancer 2: 344-358, 2011.

32. Yu JS and Cui W: Proliferation, survival and metabolism: The role of PI3K/AKT/mTOR signalling in pluripotency and cell fate determination. Development 143: 3050-3060, 2016.

33. Wang L, Zhou K, Fu Z, Yu D, Huang H, Zang X and Mo X: Brain development and Akt signaling: The crossroads of signaling pathway and neurodevelopmental diseases. J Mol Neurosci 61: 379-384, 2017.