Int J Clin Exp Med 2018;11(2):653-663
www.ijcem.com /ISSN:1940-5901/IJCEM0062136

*Original Article*

# TF-lncRNA regulation network of glioblastoma reveals specific topological features and prognostic lncRNAs

Fenghua Chen, Tao Song, Yanjin Wang, Mingyu Zhang, Zhongyi Sun, Tianxiang Huang, Jiasheng Fang, Jun Wu

*Department of Neurosurgery, Xiangya Hospital of Central South University, Changsha, China*

Received July 23, 2017; Accepted November 17, 2017; Epub February 15, 2018; Published February 28, 2018

**Abstract:** Glioblastoma multiforme (GBM) is a highly malignant tumor of human brain with poor prognosis. In recent years, long non-coding RNAs (lncRNAs) have been reported to be associated with metastasis and prognosis in GBM samples. In the transcription level, the sequence-specific transcription factors (TFs) perform critical control in lncRNAs expression. An lncRNA can be regulated by different TFs, and such co-regulation among TFs and lncRNAs defines the TF-lncRNA network that underlie complex disease. In this study, we constructed a comprehensive GBM associated TF-lncRNA regulatory network (GTLN) by integrating the sequence specific TF-lncRNA binding information and high-throughput molecular profiles of 422 TCGA GBM patients. In the GTLN, we found that TFs and lncRNAs exhibited common and specific topological features indicating the complex regulatory relationship between TFs and lncRNAs. Some hot lncRNAs were found to be involved in critical pathways and important biological processes. Further, lncRNA OSER1-AS1 was found to be associated with GBM patients' survival. The univariate and multivariate Cox regression analysis indicated that OSER1-AS1 was an independent prognostic risk factor in GBM. Overall, our analyses could provide novel insight into lncRNA-associated regulatory mechanisms and serve as helpful references for functional dissection of lncRNAs in GBM.

**Keywords:** lncRNA, transcription factors, network, glioblastoma

## Introduction

Glioblastoma multiforme (GBM) is a kind of highly malignant tumor in human brain. There are approximately 10,000 new GBM cases occur each year [1]. Despite the advances in treatment modalities, the prognosis of GBM is still poor [2]. With the development of clinical methodology, the histological diagnosis can provide helpful information for GBM treatment. However, it is insufficient for predicting survival outcomes [3]. Thus, there is an urgent need of suitable molecular biomarkers for diagnosis and prognosis of GBM patients.

In recent years, large amount of long non-coding RNAs (lncRNA), which are non-protein-coding transcripts longer than 200 bps, have been discovered in a wide range of biological functions [4]. In addition, emerging evidence reveals that lncRNAs play complex and critical roles in tumor development and pathology [5]. For example, lncRNA HOTAIR has great expression level in metastatic breast cancer patients. The inhibition of HOTAIR expression can block the metastasis progression of breast cancer [6]. The lncRNA H19 was shown to promote cancer development and invasion of glioma [7]. In addition, the expression of an oncogenic lncRNA MALAT1 has been found to be associated with metastasis and prognosis in GBM samples. A recently discovered lncRNA, HULC, has been shown to play an important role in the development of liver cancer by acting as an endogenous competing lncRNA [8]. However, these findings have provided only a limit understanding of lncRNAs in GBM. The identification and characterization of GBM related lncRNAs still remain challenge.

The acquisition of genome wide scale analysis of complex diseases has demonstrated that the process of disease pathology and tumor progression like involve in the coordinate regulation of molecular networks [9]. A previous study has constructed a GBM associated ce-

**Table 1.** Clinicopathologic properties of TCGA GBM patients (n=422)

| Characteristics | Number of patients | | | P |
|---|---|---|---|---|
| | All patients n=422 | Training set n=211 | Test set n=211 | |
| Sex | | | | 1[a] |
| Male | 262 | 131 | 131 | |
| Female | 160 | 80 | 80 | |
| Age | | | | 0.16[b] |
| Mean ± SD | 56.96±14.73 | 55.94±15.08 | 57.98±14.34 | |
| Range | 10-89 | 10-89 | 14-88 | |
| Histological type | | | | 1[a] |
| Treated primary GBM | 17 | 8 | 9 | |
| Untreated primary GBM | 405 | 203 | 202 | |
| History neoadjuvant treatment | | | | 1[a] |
| Yes | 20 | 10 | 10 | |
| No | 402 | 201 | 201 | |
| Survival (month) | | | | 0.76[b] |
| Mean ± SD | 18.24±19.30 | 17.95±17.18 | 18.53±21.25 | |
| Range | 0.10-129.37 | 0.10-96.10 | 0.13-129.37 | |
| State | | | | 0.10[a] |
| Living | 52 | 32 | 20 | |
| Death | 370 | 179 | 191 | |

[a]*P*-values were determined using Chi-square test. [b]*P*-values were determined using Student's t-test.

RNA network and identified prognostic lncRNA biomarkers [10]. In the transcription level, the sequence-specific transcription factors (TFs) perform critical control in gene expression, including lncRNAs. In particular, an lncRNA can be regulated by different TFs, and such co-regulation among TFs and lncRNAs defines the TF-lncRNA network that underlie complex disease [11]. The TF-lncRNA network could provide us a global view of all possible transcriptional interactions and further to investigate the control properties of lncRNAs. Through an analysis of the transcriptional network in brain tumors, a previous work identified the transcriptional module that controls the expression of the mesenchymal signature [12]. By constructing of TF-TF synergistic network in the progression of glioma, a previous study revealed the dynamic rewiring behavior of some TF motifs across glioma progression [13]. These results indicate that constructing and analyzing of the GBM-associated TF-lncRNA regulatory network could be necessary to understand the lncRNA functions in GBM progression.

In this work, we used a multi-step pipeline to construct a GBM associated TF-lncRNA regulatory network (GTLN). The sequence specific TF-lncRNA binding information and high-throughput molecular profiles of 422 GBM patients were integrated into the GTLN construction pipeline. Based on the systematic network analysis, we found that TFs and lncRNAs exhibited specific topological features in the GTLN. Some GBM-associated lncRNAs, such as MALAT1 and NEAT1, were found to be regulated by more TFs in the network, indicating the extensive control of TFs on disease lncRNAs. Based on functional analysis of hub and bottleneck lncRNAs nodes of GTLN, we found that these lncRNA were involved in critical pathways and important biological processes, and some of these lncRNAs were significantly associated with survival status of GBM. By building a risk model integrating of lncRNA expression and Cox regression coefficient, we found that lncRNA OSER1-AS1 could divide GBM patients into different risk groups. The univariate and multivariate Cox regression analysis indicated that OSER1-AS1 was an independent prognostic risk factor of GBM patients' survival in comparison with known clinical and pathological risk factors. Overall, these systematic analyses provided novel insight into lncRNA-associated regulatory mechanisms at transcriptional level. Both the method and predictions that were

generated in this study could serve as helpful references for future experimental and functional dissection of lncRNAs in GBM.

## Materials and methods

### *RNA expression profiles of GBM patients*

The genome-wide RNA expression levels were derived from a previous study [14], which repurposed available array-based data. Generally, the exon array data were collected from the TCGA data portal [1]. The expression values were calculated by summarizing the background-corrected intensity of all of the probes annotated to this gene [1]. Quantile normalization was performed on the expression of lncRNAs/mRNAs across patients. A combat method was used to remove potential batch effects [15]. Finally, the expression of 25496 transcripts was identified and log2 transformed in our analysis.

### *Clinical properties of patients*

The clinical and pathological data pertaining to the GBM patients were derived from the TCGA data portal. In total, 422 TCGA GBM patients with clinical follow-up information were used. The details of the information from all of the GBM patients, the training set and the validating set are summarized in **Table 1**.

### *Co-expression analysis*

To identify functional TF-lncRNA pairs, we used Pearson correlation coefficients to evaluate the co-expression relationship based on the expression level between each potential TF-lncRNA pairs as follows:

$$\rho_{X,Y} = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y}, \quad (1)$$

cov(*X*, *Y*) is the covariance of variables X and Y. $\sigma_X$ and $\sigma_Y$ are the standard deviations for X and Y, respectively. The threshold was set to an FDR<0.01.

### *Construction of the risk score model*

To evaluate the risk factors that predict the survival of the GBM patients, the patients were randomly assigned to a training data set or a validating data set (**Table 1**). The two subsets were required to have the same size and have no significant difference in clinical factors (Chi-square test or Student's t-test, P>0.05). Univariate Cox regression analysis was used to evaluate the association between survival and the expression level of each candidate. After the univariate Cox regression analysis, a risk score formula was constructed by integrating both the strength and positive/negative association between each candidate and survival. The risk score for each patient was calculated according to the linear combination of the expression values weighted by the regression coefficient from the univariate Cox regression analysis:

$$RiskScore = \sum_{i=1}^{n} r_i Exp(i), \quad (2)$$

in which $r_i$ is the Cox regression coefficient of candidate i from the training set, and n is the number of testing candidates. *Exp(i)* is the expression value of candidate i in a corresponding patient. The median risk score was used as the cut-off to classify the training dataset into the high- and low-risk groups. The patients in the high-risk group were expected to have poor survival outcomes. Conversely, the patients in the low-risk group were expected to have high survival outcomes. This model and cut-off point was further applied to the validating set to divide the patients into high- and low-risk groups.

### *Network illustration and topological analysis*

We used Cytoscape software (v3.1.1) to construct and illustrate the ceRNA network. Several topological properties, such as the node degree, BC, CC and NC, were analyzed by the built-in Network Analyzer tool in Cytoscape.

### *Statistical analysis*

All analyses were performed based on R 3.1.0 software. The Kaplan-Meier survival analysis was performed for different groups of patients, and statistical significance was assessed using the log-rank test (P<0.05). Pearson correlation coefficient was used to evaluate the co-expression relationship based on the expression level between each potential TF-lncRNA pairs (FDR<0.01). The GBM patients were randomly assigned to a training data set or a validating data set (**Table 1**). In two groups of GBM

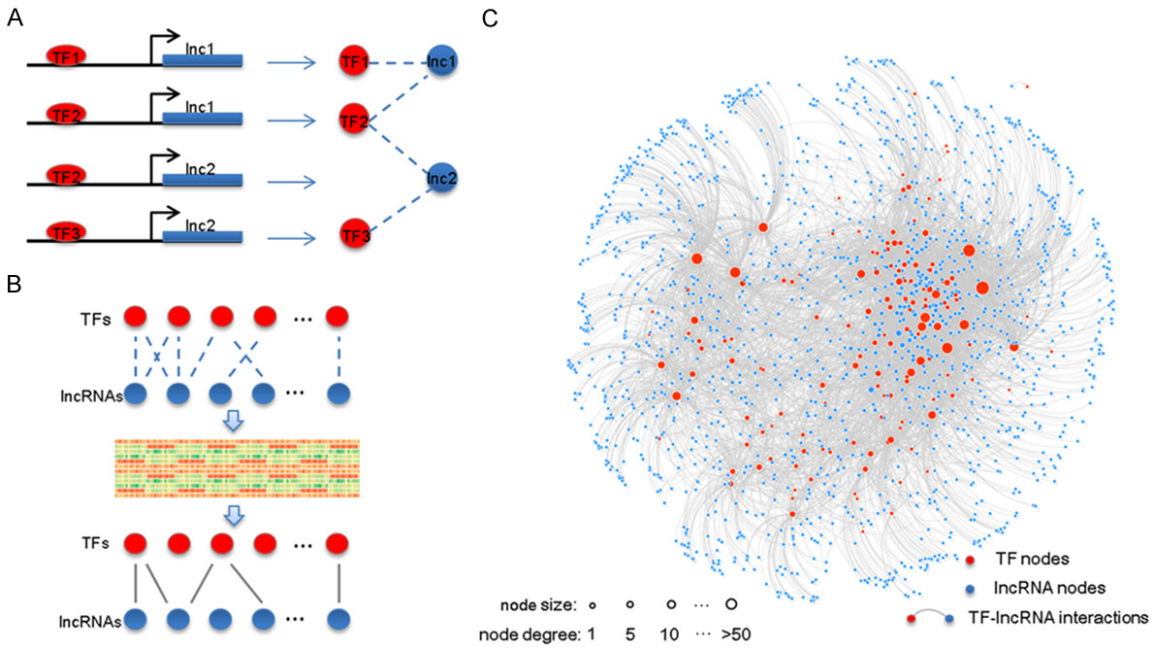


**Figure 1.** Construction of TF-lncRNA interaction network of GBM. A. Identification of candidate TF-lncRNA interactions through sequence-based dataset. B. Inferring the regulatory relationships between TFs and lncRNAs which are specifically co-expressed in GBM patients. C. A global view of the GTLN. In GTLN, TFs and lncRNAs are represented as red and blue nodes, respectively. An edge represents a TF-lncRNA regulation in GBM.

patients, significant difference of several clinical factors, including sex, histological type, history neoadjuvant treatment and survival status was assessed using Chi-square test ($P<0.05$).

Age and survival days were evaluated by Student's t-test ($P<0.05$).

## Results

### *Systematic identification of TF-lncRNA interactions and construction of GTLN*

To evaluate the landscape of TFs regulation on lncRNAs in GBM, we used a multi-step approach to identify functional TF-lncRNA interactions by integrating sequence specific TF-lncRNA binding and TF-lncRNA co-expression information. We derived TF-lncRNA binding datasets from SNP@lincTFBS database [16], which identifies transcription factor binding sites of lncRNA using genome wide CHIP-Seq data (**Figure 1A**). These dataset has been used to identify functional lncRNA-mediated transcription feed-forward loop in different cancers [17]. However, evidence of theoretical TF binding does not directly imply the active transcriptional regulation of lncRNA. Inferring the regulatory relationships between TFs and lncRNAs which are specifically co-expressed in GBM patients can be performed based on the expression data. We calculated the Pearson correlation coefficient for each candidate TF-lncRNA pair identified above (**Figure 1B**). Finally, significantly co-expressed TF-lncRNA interactions were used to construct the GTLN. Further, the GTLN was graphically modeled while TFs and lncRNAs were viewed as nodes and the transcription regulations were viewed as interacting lines (**Figure 1C**). The GTLN contained 145 TFs and 1489 lncRNAs, and 6071 TF-lncRNA interactions.

### *Common and specific topological properties of GTLN*

The GTLN could serve a global view of all possible TF-lncRNA interactions in which we can investigate the regulatory properties of TFs and lncRNAs. An analysis of the topological properties of the TF-lncRNA regulatory network revealed come common features. Most TFs and lncRNAs are connected and form a large connecting subnetwork. We found that approximately 94% of the TFs in the GTLN regulate at least two lncRNAs, and about 62% of lnc-

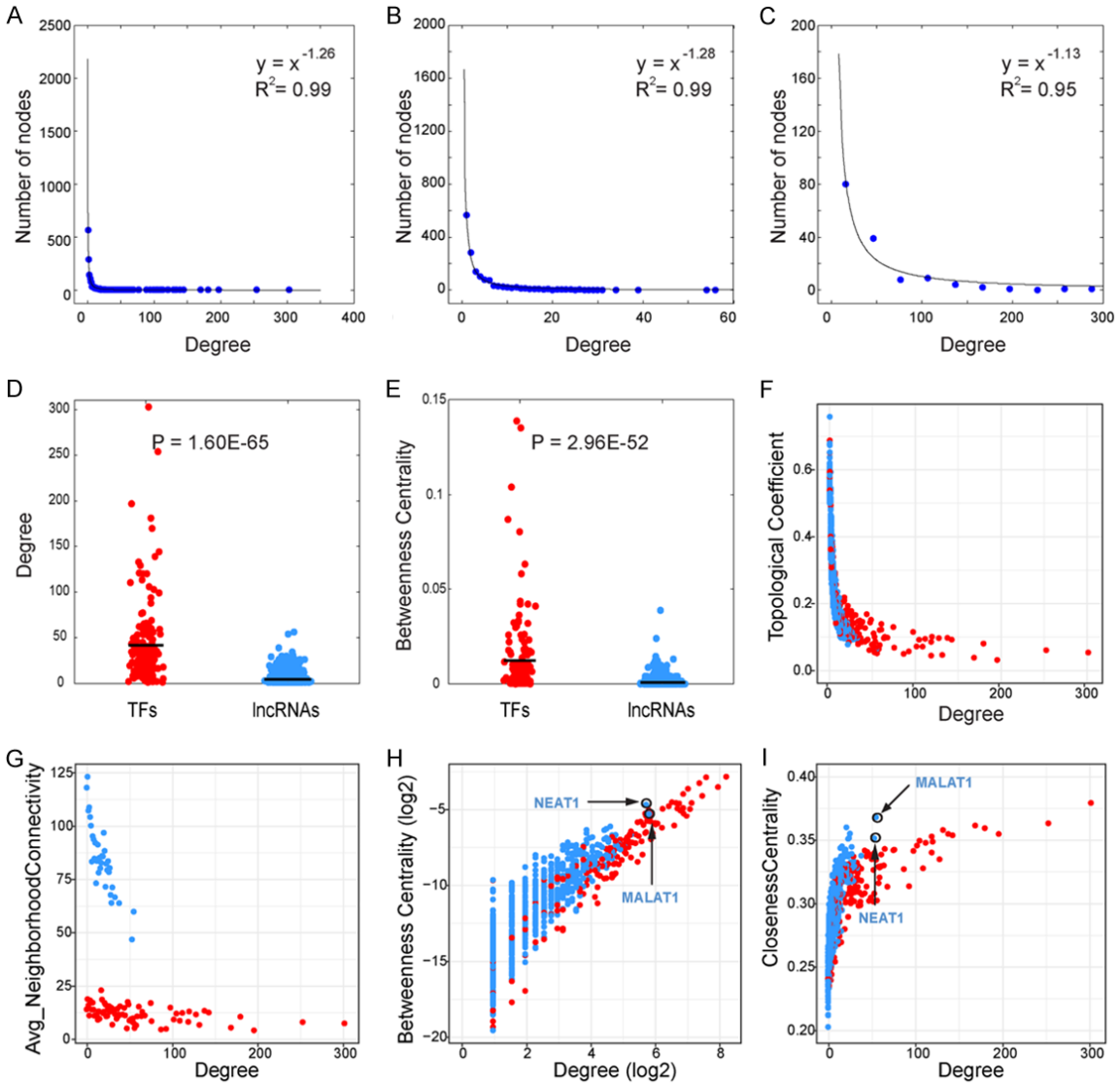


**Figure 2.** Comprehensive topological analysis of GTLN. The degree distribution of the entire network (A), lncRNA nodes (B) and the TF nodes (C) reveals power-law distributions of GTLN network. (D) TF nodes had higher degrees than lncRNA nodes. (E) TF nodes had higher BCs than lncRNA nodes. (F) The topological coefficient distribution along with degree of nodes in GTLN. (G-I) In GTLN, TF nodes and lncRNA nodes exhibit different average NC, BC and CC with the same node degrees. Some known GBM-associated lncRNAs were involved in GTLN and tend to be hub and bottleneck nodes.

RNAs are co-regulated by two of more TFs. These results indicated a complicated combination of TFs regulation on lncRNA targets. We evaluated the degree distribution of the entire network (**Figure 2A**, R2=0.99), lncRNA nodes (**Figure 2B**, $R^2$=0.99) and the TF nodes (**Figure 2C**, $R^2$=0.94). The degree property of the nodes reveals power-law distributions, which indicate that the GBM-associated TF-lncRNA interaction network was a scale-free network like most of biological networks [18]. Together with degree, we considered the betweenness centrality (BC) property and compare these features between TFs and lncRNAs (**Figure 2D**, **2E**). We found that TF nodes had more degrees (P=1.60E-65) and BCs (P=2.96E-52) in comparison with lncRNA nodes, indicating the controlling force of TFs on lncRNAs expression. The constant decrease in the topological coefficient as the degree increase for each node indicates the GTLN have a hierarchical modularity (**Figure 2F**) [19]. Further, we calculated neighbor-

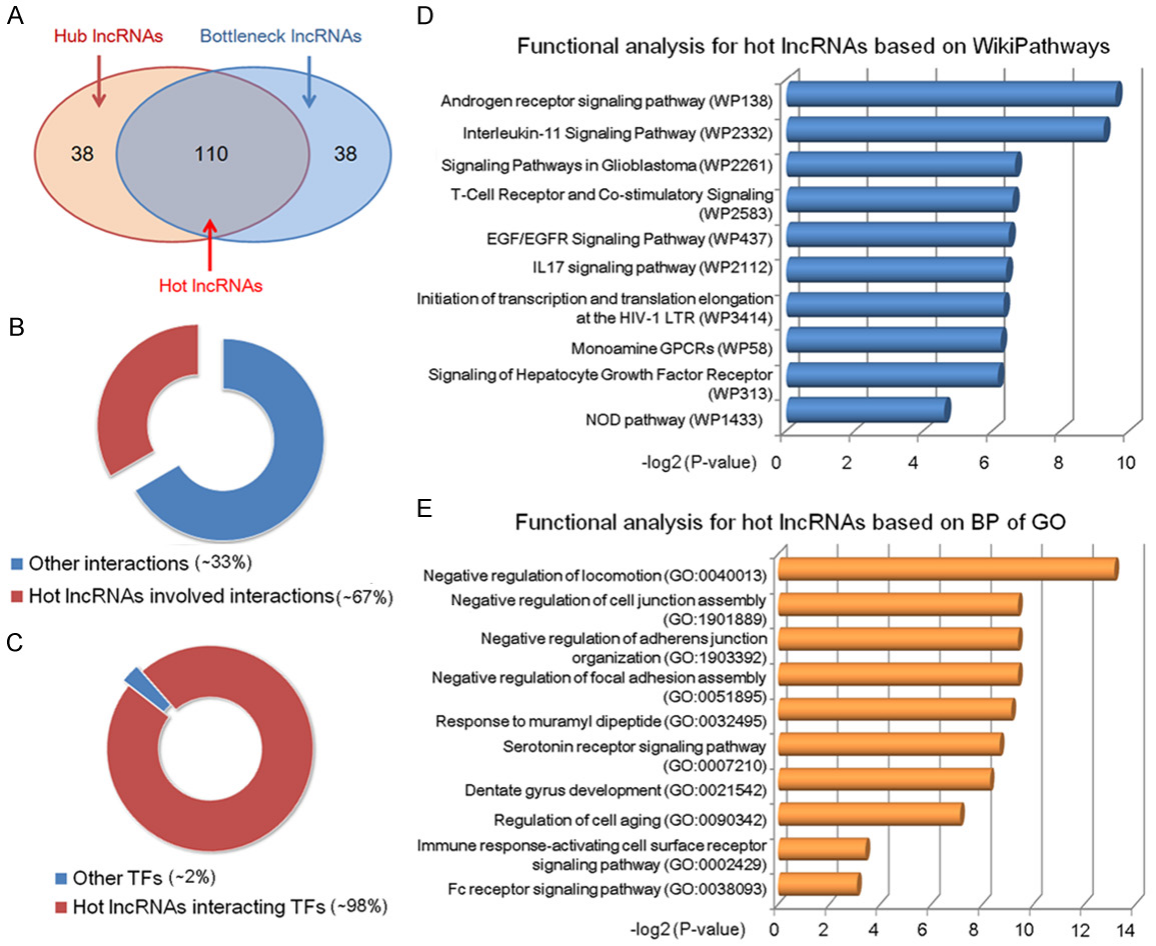


**Figure 3.** Functional analysis of the hot lncRNAs in GTLN. A. 110 hot lncRNAs were found in the overlap of hub and bottneck sets. B. The hot lncRNAs were involved in 33% of the interactions. C. The hot lncRNAs intacting with 98% of the TFs in the GTLN. D. Functional analysis of hot lncRNAs based on Wiki pathways. E. Functional analysis of hot lncRNAs based on BP of GO.

hood connectivity (NC) and closeness centrality (CC) of all the nodes within. The NC distribution provides the average of the neighborhood connective status of TFs and lncRNAs. The CC is to evaluate the degree of a node to be central in a given network, by taking a reciprocal of an average shortest path length to all the nodes. We found that TF nodes and lncRNA nodes exhibit different average NC, BC and CC with the same node degrees (**Figure 3G-I**). These specific properties reveal the different regulating functions of TFs and lncRNAs in GTLN.

Some well-known GBM-associated lncRNAs are involved in the GTLN. For example, the enhanced expression of lncRNA MALAT1 confers a potent poor therapeutic efficacy [20]. The knockdown of MALAT1 can reduced GBM cell migration indicating the inhibition of MALAT1 levels could be a future direction to develop a novel therapeutic strategy of GBM [21]. The lncRNA NEAT1 has been found to promote oncogenesis in various tumors. A recent study has found that NEAT1 was upregulated in GBM and promoted oncogenesis by downregulating let-7e expression [22]. These two GBM-associated lncRNAs exhibit specific topological properties in GTLN. With the same degree, MALAT1 and NEAT1 had high BC and CC value than other lncRNAs (**Figure 2H**, **2I**). These specific properties could help us to identify novel disease-associated lncRNAs on the context of GTLN.

*Functional analysis of hot lncRNAs in GTLN*

Based on the above observations, we found that NEAT1 was rank as top 1 in BC and top 2 in

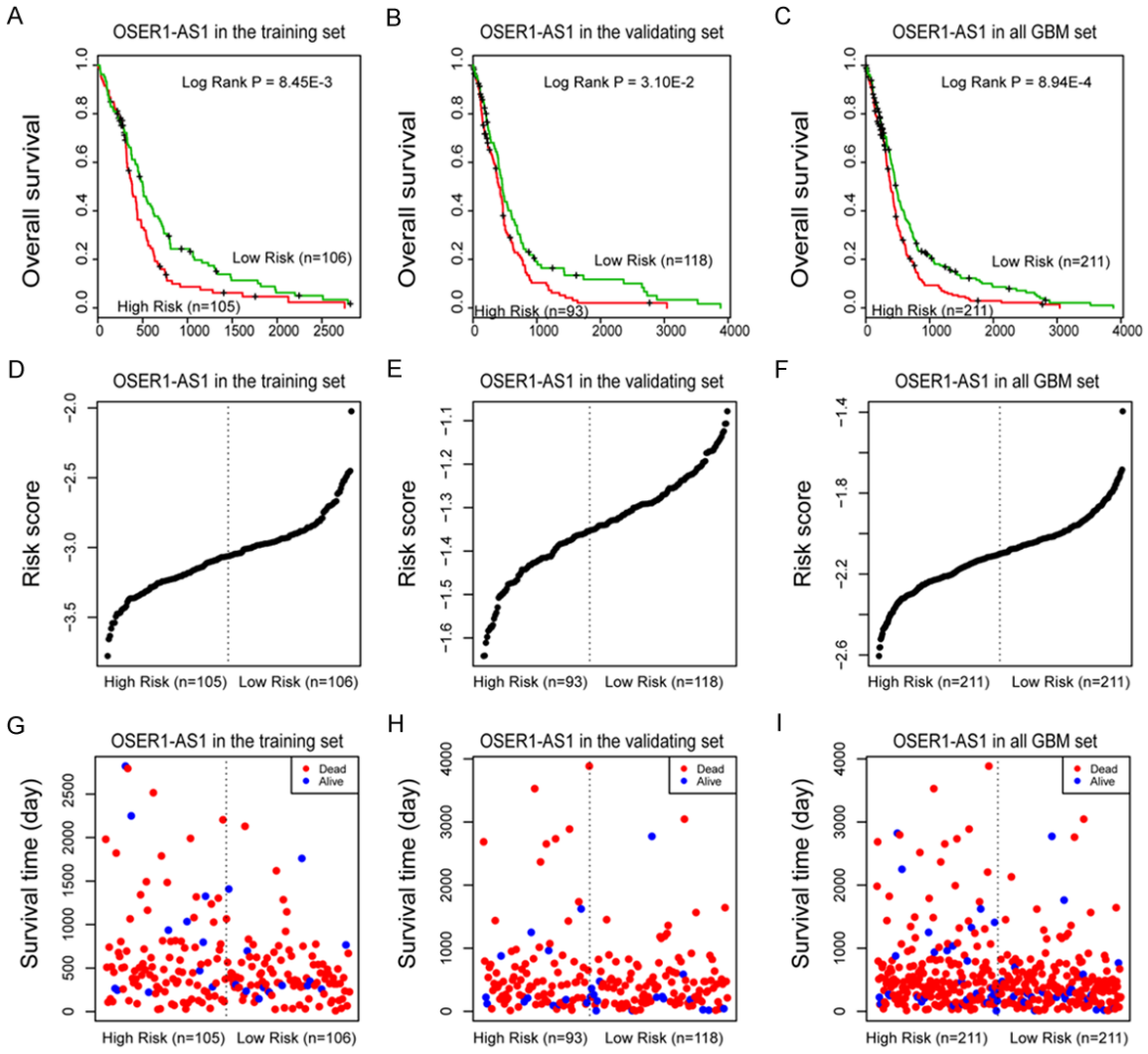


**Figure 4.** Comprehensive survival analysis of lncRNA OSER1-AS1 in GBM patients. A. OSER1-AS1 could divided the training GBM patients into two different risk groups (P=8.45E-3). B. OSER1-AS1 could divided the validating GBM patients into two different risk groups (P=3.10E-2). C. All 422 GBM patients could be divided into two different risk groups by OSER-AS1 (P=8.94E-4). D-F. The risk score distribution of different groups of patients in training, validating and all GBM patients. G-I. Individual plot of patients' survival time of different groups in train, validating and all GBM patients. The grey dashed lines represent the risk score thresholds in survival analysis.

degree, while MALAT1 was ranked as top 2 in BC and top 1 in degree among lncRNAs in GTLN. These lncRNAs were hub and bottleneck nodes in the biological network. In general, a higher degree indicates that the node is a hub that participating in more network interactions. A higher BC implies that the node is a bottleneck that acting as bridges connecting different network modules. In previous studies, hubs and bottlenecks nodes were typically defined as the top 10%-20% of the nodes ranked by degree or BC [23]. In this work, we used the top 10% as a threshold to define hub and bottleneck lncRNAs in GTLN. We found a set of 148 lncRNAs as hubs and another set of 148 lncRNAs as bottlenecks. There are 110 lncRNAs were found in the overlap of hub and bottleneck sets (**Figure 3A**). These lncRNA nodes were defined as hot lncRNAs. The hot lncRNAs were involved in 33% of the interactions and regulated by 98% of the TFs in the GTLN (**Figure 3B**, **3C**), indicating that the major transcriptional regulation and crosstalk between TFs were undertaken by these hot lncRNAs. Next, we explored the functions of hot lncRNAs by using the Enrich web based tool

**Table 2.** Univariate and multivariate analysis of clinicopathological factors and OSER1-AS1 in GBM

| Variables | Univariate analysis | | | Multivariate analysis | | |
|---|---|---|---|---|---|---|
| | HR (95% CI) | Coefficient | P | HR (95% CI) | Coefficient | P |
| Sex | 1.186 (0.956-1.469) | 0.171 | 0.12 | 1.064 (0.856-1.319) | 0.062 | 0.57 |
| Age | 1.030 (1.023-1.039) | 0.030 | 7.59E-14 | 1.017 (1.0041-1.029) | 0.026 | 3.63E-11 |
| History neoadjuvant treatment | 0.456 (0.285-0.728) | -0.785 | 1.02E-03 | 1.061 (0.8083-1.392) | -0.479 | 0.25 |
| Histological type | 0.510 (0.308-0.846) | -0.673 | 9.08E-03 | 1.202 (0.8937-1.617) | -0.005 | 0.99 |
| OSER1-AS1 | 0.747 (0.627-0.889) | -0.292 | 1.03E-03 | 1.737 (1.0808-2.790) | -0.257 | 4.15E-03 |

[24, 25], which performs a comprehensive gene set enrichment analysis based on different functional context such as Gene Ontology (GO), Wiki Pathway, KEGG Pathway and etc. We found that these hot lncRNAs were associated with some critical pathways in GBM. For example, the EGFR Signaling Pathway (**Figure 3D**) is an oncogenic pathway of GBM [26]. This pathway activates SREBP-1 and its regulated lipid synthesis and uptake pathways via upregulation of SCAP, to promote rapid GBM growth [27]. Signaling Pathways of Glioblastoma also can be found in the function list (**Figure 4D**). Based on the context of GO, some important biological processes (BP), which is associate with tumor pathology, are significantly enriched. For example, the regulation of adherens junction (GO:1903392) and focal adhesion assembly (GO:0051895) have been reported as tumor inhibiting functions. These functions were associate with clinical outcome of patients with malignant gliomas and may serve as a promising tumor suppressor-related processes [28]. Results of functional analysis based on other context were listed in Tables S1, S2, S3.

*Identification of prognostic lncRNAs in GBM*

Based on the specific topologies and critical functions of the hot lncRNAs in the GTLN, we hypothesized that these hot lncRNAs might be involved in the pathological processes and risk factors of GBM. Cox regression analysis was performed to investigate whether these hot lncRNAs were prognostic factors in GBM. Overall, among the 110 hot lncRNAs, 22 were found to be significantly associated with the survival of 422 GBM patients (Table S4). To test whether these lncRNAs were prognostic factors for GBM, a risk model was constructed (Materials and Methods). The 422 GBM patients were randomly assigned into two groups that were used as training (n=211) and validating (n=211) datasets (**Table 1**). There were no significant differences between the two groups of patients in the clinical factors (P>0.05).

Kaplan-Meier survival analysis of the training GBM patients revealed that an lncRNA named OSER1-AS1, which is a protective factor (Cox coefficient =-0.29, P=1.03E-3), could divided the training GBM patients into two different risk groups (**Figure 4A**, P=8.45E-3). Next, we investigated OSER1-AS1 in the validating dataset using the same risk score threshold as in the training set. Based on this strategy, the validating patients were divided high- and low-risk groups (**Figure 4B**, P=3.10E-2). Further, we used OSER1-AS1 as a prognostic biomarker to divide all 422 GBM patients. Same as training and validating dataset, all 422 GBM patients could be divided into two different risk groups (**Figure 4C**, P=8.94E-4). The high-risk group consisted of patients with high risk scores and had a lower survival times (**Figure 4D-I**). These results indicated that OSER1-AS1 could be used as a potential prognostic factor for GBM.

*LncRNA OSER1-AS1 is an independent prognosis biomarker*

To further test whether the lncRNA OSER1-AS1 is an independent predictor of GBM patients survival, the prognostic association between this lncRNA and some other known clinical and pathological risk factors in GBM progression were evaluated by univariate and multivariate analyses. Several clinicopathologic factors, such as age, sex, histological type and history neoadjuvant treatment, were considered. As expected, in addition to patient age, which is already a known risk factor, OSER1-AS1 was a significant risk factor for survival in a univariate analysis (**Table 2**, P=1.03E-3). A multivariate analysis further revealed that the lncRNA OSER1-AS1 remained an independent prognostic risk factor for survival (**Table 2**, P=4.15E-3).

## Discussion

In recent years, large amount of lncRNAs have been discovered in a wide range of biological functions [4]. The emerging evidence reveals that lncRNAs play important roles in tumor development and pathology [5]. The acquisition of genome wide scale analysis demonstrated that the process of disease pathology and tumor progression likely to be involve in the coordinate regulation of biological regulating networks [9]. A previous work has constructed a GBM associated ceRNA network and identified prognostic lncRNA biomarkers [10]. However, these findings have provided only a limit understanding of lncRNAs in GBM, which is a kind of highly malignant tumor in human brain. The identification and characterization of GBM related lncRNAs still remain challenge.

In particular, lncRNAs can be regulated by different TFs, and such co-regulation among TFs and lncRNAs defines the TF-lncRNA network across different diseases. The TF-lncRNA network could provide us a global view of all possible transcriptional interactions. By constructing of TF-TF synerstic network in the progression of glioma, a study revealed the dynamic rewiring behavior of some TF motifs across glioma progression [13]. These results indicate that constructing and analyzing of the GBM-associated TF-lncRNA regulatory network could be necessary to understand the lncRNA functions in GBM progression.

In this work, we integrated sequence specific TF-lncRNA binding information and high-throughput molecular profiles into a computational pipeline to construct a GBM associated TF-lncRNA interaction network. 422 GBM samples from TCGA database have been analyzed. Through network analysis, we found that TFs and lncRNAs exhibited specific topological features in the GTLN. Some GBM-associated lncRNAs, such as NEAT1 and MALAT1, were found to be regulated by more TFs than other lncRNAs, indicating the strongly control of disease lncRNAs by TFs. Based on functional analysis of hub and bottleneck nodes of GTLN, we found that these lncRNA were involved in important pathways and biological processes, and some of these lncRNAs were significantly associated with survival status. By building a risk model integrating of expression and Cox regression coefficient, we found that lncRNA OSER1-AS1 could divide GBM patients into different risk groups. The univariate and multivariate Cox regression analysis indicated that OSER1-AS1 was an independent prognostic risk factor of GBM patients' survival in comparison with known clinical and pathological risk factors. In conclusion, our analyses provided novel insight into lncRNA-associated regulatory mechanisms at transcriptional level. Both the method and predictions could serve as helpful references for future experimental and further functional dissection of lncRNAs.

## Disclosure of conflict of interest

None.

**Address correspondence to:** Jun Wu, Department of Neurosurgery, Xiangya Hospital of Central South University, Changsha, China. E-mail: wujundc@163.com

## References

[1] Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature 2008; 455: 1061-1068.

[2] Stupp R, Mason WP, van den Bent MJ, Weller M, Fisher B, Taphoorn MJ, Belanger K, Brandes AA, Marosi C, Bogdahn U, Curschmann J, Janzer RC, Ludwin SK, Gorlia T, Allgeier A, Lacombe D, Cairncross JG, Eisenhauer E, Mirimanoff RO; European Organisation for Research and Treatment of Cancer Brain Tumor and Radiotherapy Groups; National Cancer Institute of Canada Clinical Trials Group. Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. N Engl J Med 2005; 352: 987-996.

[3] Verhaak RG, Hoadley KA, Purdom E, Wang V, Qi Y, Wilkerson MD, Miller CR, Ding L, Golub T, Mesirov JP, Alexe G, Lawrence M, O'Kelly M, Tamayo P, Weir BA, Gabriel S, Winckler W, Gupta S, Jakkula L, Feiler HS, Hodgson JG, James CD, Sarkaria JN, Brennan C, Kahn A, Spellman PT, Wilson RK, Speed TP, Gray JW, Meyerson M, Getz G, Perou CM, Hayes DN; Cancer Genome Atlas Research Network. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. Cancer Cell 2010; 17: 98-110.

[4] Ponting CP, Oliver PL and Reik W. Evolution and functions of long noncoding RNAs. Cell 2009; 136: 629-641.

[5] Wilusz JE, Sunwoo H and Spector DL. Long noncoding RNAs: functional surprises from the RNA world. Genes Dev 2009; 23: 1494-1504.
[6] Gupta RA, Shah N, Wang KC, Kim J, Horlings HM, Wong DJ, Tsai MC, Hung T, Argani P, Rinn JL, Wang Y, Brzoska P, Kong B, Li R, West RB, van de Vijver MJ, Sukumar S and Chang HY. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature 2010; 464: 1071-1076.
[7] Shi Y, Wang Y, Luan W, Wang P, Tao T, Zhang J, Qian J, Liu N and You Y. Long non-coding RNA H19 promotes glioma cell invasion by deriving miR-675. PLoS One 2014; 9: e86295.
[8] Wang J, Liu X, Wu H, Ni P, Gu Z, Qiao Y, Chen N, Sun F and Fan Q. CREB up-regulates long non-coding RNA, HULC expression through interaction with microRNA-372 in liver cancer. Nucleic Acids Res 2010; 38: 5366-5383.
[9] Wang P, Ning S, Zhang Y, Li R, Ye J, Zhao Z, Zhi H, Wang T, Guo Z and Li X. Identification of lncRNA-associated competing triplets reveals global patterns and prognostic markers for cancer. Nucleic Acids Res 2015; 43: 3478-3489.
[10] Cao Y, Wang P, Ning S, Xiao W, Xiao B and Li X. Identification of prognostic biomarkers in glioblastoma using a long non-coding RNA-mediated, competitive endogenous RNA network. Oncotarget 2016; 7: 41737-41747.
[11] Tsankov AM, Gu H, Akopian V, Ziller MJ, Donaghey J, Amit I, Gnirke A and Meissner A. Transcription factor binding dynamics during human ES cell differentiation. Nature 2015; 518: 344-349.
[12] Carro MS, Lim WK, Alvarez MJ, Bollo RJ, Zhao X, Snyder EY, Sulman EP, Anne SL, Doetsch F, Colman H, Lasorella A, Aldape K, Califano A and Iavarone A. The transcriptional network for mesenchymal transformation of brain tumours. Nature 2010; 463: 318-325.
[13] Li Y, Shao T, Jiang C, Bai J, Wang Z, Zhang J, Zhang L, Zhao Z, Xu J and Li X. Construction and analysis of dynamic transcription factor regulatory networks in the progression of glioma. Sci Rep 2015; 5: 15953.
[14] Du Z, Fei T, Verhaak RG, Su Z, Zhang Y, Brown M, Chen Y and Liu XS. Integrative genomic analyses reveal clinically relevant long noncoding RNAs in human cancer. Nat Struct Mol Biol 2013; 20: 908-913.
[15] Johnson WE, Li C and Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics 2007; 8: 118-127.
[16] Ning S, Zhao Z, Ye J, Wang P, Zhi H, Li R, Wang T, Wang J, Wang L and Li X. SNP@lincTFBS: an integrated database of polymorphisms in human LincRNA transcription factor binding sites. PLoS One 2014; 9: e103851.
[17] Ning S, Gao Y, Wang P, Li X, Zhi H, Zhang Y, Liu Y, Zhang J, Guo M, Han D and Li X. Construction of a lncRNA-mediated feed-forward loop network reveals global topological features and prognostic motifs in human cancers. Oncotarget 2016; 7: 45937-45947.
[18] Nacher JC and Akutsu T. Recent progress on the analysis of power-law features in complex cellular networks. Cell Biochem Biophys 2007; 49: 37-47.
[19] Simos T, Georgopoulou U, Thyphronitis G, Koskinas J and Papaloukas C. Analysis of protein interaction networks for the detection of candidate hepatitis B and C biomarkers. IEEE J Biomed Health Inform 2015; 19: 181-189.
[20] Chen W, Xu XK, Li JL, Kong KK, Li H, Chen C, He J, Wang F, Li P, Ge XS and Li FC. MALAT1 is a prognostic factor in glioblastoma multiforme and induces chemoresistance to temozolomide through suppressing miR-203 and promoting thymidylate synthase expression. Oncotarget 2017; 8: 22783-22799.
[21] Vassallo I, Zinn P, Lai M, Rajakannu P, Hamou MF and Hegi ME. WIF1 re-expression in glioblastoma inhibits migration through attenuation of non-canonical WNT signaling by down-regulating the lncRNA MALAT1. Oncogene 2016; 35: 12-21.
[22] Gong W, Zheng J, Liu X, Ma J, Liu Y and Xue Y. Knockdown of NEAT1 restrained the malignant progression of glioma stem cells by activating microRNA let-7e. Oncotarget 2016; 7: 62208-62223.
[23] Xu J, Li Y, Lu J, Pan T, Ding N, Wang Z, Shao T, Zhang J, Wang L and Li X. The mRNA related ceRNA-ceRNA landscape and significance across 20 major cancer types. Nucleic Acids Res 2015; 43: 8169-8182.
[24] Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, Clark NR and Ma'ayan A. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics 2013; 14: 128.
[25] Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A, McDermott MG, Monteiro CD, Gundersen GW and Ma'ayan A. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res 2016; 44: W90-97.
[26] Ru P, Hu P, Geng F, Mo X, Cheng C, Yoo JY, Cheng X, Wu X, Guo JY, Nakano I, Lefai E, Kaur B, Chakravarti A and Guo D. Feedback loop regulation of SCAP/SREBP-1 by miR-29 modulates EGFR signaling-driven glioblastoma growth. Cell Rep 2016; 16: 1527-1535.

[27] Cheng C, Ru P, Geng F, Liu J, Yoo JY, Wu X, Cheng X, Euthine V, Hu P, Guo JY, Lefai E, Kaur B, Nohturfft A, Ma J, Chakravarti A and Guo D. Glucose-mediated N-glycosylation of SCAP is essential for SREBP-1 activation and tumor growth. Cancer Cell 2015; 28: 569-581.

[28] Zeng L, Fee BE, Rivas MV, Lin J and Adamson DC. Adherens junctional associated protein-1: a novel 1p36 tumor suppressor candidate in gliomas (Review). Int J Oncol 2014; 45: 13-17.

**Table S1.** Functional analysis of hot lncRNAs based on KEGG context

| Functions | *P*-value |
|---|---|
| Hepatitis B Homo sapiens hsa05161 | 7.62E-03 |
| Hepatitis C Homo sapiens hsa05160 | 5.51E-03 |
| Chemokine signaling pathway Homo sapiens hsa04062 | 1.76E-02 |
| Serotonergic synapse Homo sapiens hsa04726 | 2.17E-02 |
| VEGF signaling pathway Homo sapiens hsa04370 | 4.16E-02 |

**Table S2.** Functional analysis of hot lncRNAs based on Molecular Function of GO

| Functions | *P*-value |
|---|---|
| Amine binding (GO:0043176) | 9.78E-04 |
| Serotonin binding (GO:0051378) | 7.63E-04 |
| Serotonin receptor activity (GO:0004993) | 2.43E-03 |
| RNA polymerase II core promoter proximal region sequence-specific DNA binding transcription factor activity (GO:0000982) | 4.64E-03 |
| RNA polymerase II core promoter proximal region sequence-specific DNA binding transcription factor activity involved in negative regulation of transcription (GO:0001078) | 4.56E-03 |
| Core promoter proximal region DNA binding (GO:0001159) | 1.49E-02 |
| Core promoter proximal region sequence-specific DNA binding (GO:0000987) | 1.44E-02 |
| RNA polymerase II core promoter proximal region sequence-specific DNA binding (GO:0000978) | 1.21E-02 |
| RNA polymerase II transcription regulatory region sequence-specific DNA binding transcription factor activity involved in negative regulation of transcription (GO:0001227) | 1.15E-02 |
| Delayed rectifier potassium channel activity (GO:0005251) | 1.56E-02 |

**Table S3.** Functional analysis of hot lncRNAs based on Cellular Component of GO

| Functions | *P*-value |
|---|---|
| Proteinaceous extracellular matrix (GO:0005578) | 3.84E-02 |
| Unconventional myosin complex (GO:0016461) | 4.16E-02 |

**Table S4.** Univariate Cox regression analysis of the hot lncRNAs in the GTLN

| LncRNAs | Ensembl ID | Univariate Cox analysis | | |
|---|---|---|---|---|
| | | HR (95% CI) | Coefficient | *P*-value |
| AC004895.4 | ENSG00000231704 | 0.862 (0.756-0.982) | -0.15 | 2.57E-02 |
| LINC01089 | ENSG00000212694 | 0.790 (0.679-0.918) | -0.24 | 2.10E-03 |
| AC092835.2 | ENSG00000233757 | 0.827 (0.703-0.974) | -0.19 | 2.25E-02 |
| AC093627.9 | ENSG00000242474 | 0.736 (0.563-0.96) | -0.31 | 2.40E-02 |
| AP001258.4 | ENSG00000245571 | 1.401 (1.008-1.946) | 0.34 | 4.45E-02 |
| AP001432.14 | ENSG00000242553 | 0.872 (0.768-0.99) | -0.14 | 3.41E-02 |
| MIR9-3HG | ENSG00000255571 | 0.897 (0.82-0.981) | -0.11 | 1.78E-02 |
| MIR155HG | ENSG00000234883 | 1.141 (1.045-1.246) | 0.13 | 3.17E-03 |
| NEAT1 | ENSG00000245532 | 1.170 (1.043-1.313) | 0.16 | 7.34E-03 |
| OSER1-AS1 | ENSG00000223891 | 0.747 (0.627-0.889) | -0.29 | 1.03E-03 |
| RP11-1055B8.3 | ENSG00000262223 | 0.575 (0.434-0.762) | -0.55 | 1.20E-04 |
| RP11-120D5.1 | ENSG00000234129 | 0.825 (0.705-0.964) | -0.19 | 1.57E-02 |
| LINC02175 | ENSG00000262155 | 0.798 (0.645-0.988) | -0.23 | 3.82E-02 |
| SNHG19 | ENSG00000260260 | 0.820 (0.704-0.955) | -0.20 | 1.09E-02 |
| RP11-439L18.1 | ENSG00000232618 | 0.768 (0.603-0.978) | -0.26 | 3.23E-02 |
| LCMT1-AS1 | ENSG00000260448 | 0.609 (0.488-0.76) | -0.50 | 1.18E-05 |
| RP11-451B8.1 | ENSG00000239572 | 1.147 (1.019-1.291) | 0.14 | 2.27E-02 |
| RP11-480A16.1 | ENSG00000260261 | 0.897 (0.807-0.996) | -0.11 | 4.25E-02 |
| RP11-543C4.1 | ENSG00000247970 | 0.742 (0.589-0.936) | -0.30 | 1.18E-02 |
| RP11-80H5.7 | ENSG00000240996 | 0.880 (0.794-0.974) | -0.13 | 1.40E-02 |
| RP5-888M10.2 | ENSG00000229484 | 0.710 (0.571-0.884) | -0.34 | 2.18E-03 |
| TMEM72-AS1 | ENSG00000224812 | 0.662 (0.498-0.882) | -0.41 | 4.79E-03 |