Identification of an EMT-related lncRNA signature in glioma

Chuming Tao

Department of Neurosurgery, The Second Affiliated Hospital of Nanchang University, Nanchang, China.
Scientific Research Center, East China Institute of Digital Medical Engineering, Shangrao, China

Qing Hu

Department of Neurosurgery, The Second Affiliated Hospital of Nanchang University, Nanchang, China.
Scientific Research Center, East China Institute of Digital Medical Engineering, Shangrao, China

Haitao Luo

Department of Neurosurgery, The Second Affiliated Hospital of Nanchang University, Nanchang, China.
Scientific Research Center, East China Institute of Digital Medical Engineering, Shangrao, China

Peng Wang

Department of Neurosurgery, The Second Affiliated Hospital of Nanchang University, Nanchang, China

Zewei Tu

Department of Neurosurgery, The Second Affiliated Hospital of Nanchang University, Nanchang, China

Hua Guo

Department of Neurosurgery, The Second Affiliated Hospital of Nanchang University, Nanchang, China

Kai Huang

Department of Neurosurgery, The Second Affiliated Hospital of Nanchang University, Nanchang, China

Xingen Zhu (✉ zxg2008vip@163.com)

The Second Affiliated Hospital of Nanchang University <https://orcid.org/0000-0002-8556-0941>

Research

Keywords: epigenetics, epithelial-mesenchymal transition, glioma, long noncoding RNA, RNA sequencing, prognostic signature

Posted Date: May 17th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-28627/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background Epithelial-mesenchymal transition (EMT) has been implicated in the invasion and progression of gliomas, the present study investigated EMT-related long noncoding (lnc)RNAs in gliomas.

Methods Candidate lncRNAs screened from a glioma dataset in The Cancer Genome Atlas were used to construct EMT-related lncRNA coexpression networks in order to identify EMT-related lncRNAs; these were validated using the Chinese Glioma Genome Atlas (CGGA). Gene set enrichment analysis and principal component analysis (PCA) were used to annotate lncRNA functions. We established an EMT-related lncRNA signature composed of 9 lncRNAs (HAR1A, LINC00641, LINC00900, MIR210HG, MIR22HG, PVT1, SLC25A21-AS1, SNAI3-AS1, and SNHG18) that could distinguish between low- and high-risk glioma patients. The functions of the identified lncRNAs were evaluated by constructing a competing endogenous RNA network.

Results The low-risk group had longer overall survival (OS) than the high-risk group ($P < 0.0001$). High-risk patients also had no deletion on chromosome arms 1p and/or 19q, expressed wild-type isocitrate dehydrogenase, and had higher World Health Organization (WHO) tumor grade. The signature was an independent factor significantly associated with OS ($P = 0.041$, hazard ratio = 1.806). These findings were validated in the CGGA dataset. PCA also revealed differences in EMT status between low- and high-risk patients. Competing endogenous RNA network showed that complex lncRNA–microRNA–mRNA interactions potentially contributing to EMT progression in glioma.

Conclusion Our results demonstrate that the EMT-related 9-lncRNA signature has prognostic value for glioma patients.

Introduction

Gliomas are the most common type of malignant brain tumor in adults[1]. Especially, the glioblastoma (WHO grade IV) is the most invasive and aggressive glioma subtype with the highest mortality rate [2; 3]. It is said that epithelial-mesenchymal transition (EMT) is closely related to glioma malignancies[4; 5], and we have known that lncRNAs orchestrate multiple cellular processes by modulating EMT in a variety of tumors cells[6]. Even though the detailed mechanisms of lncRNAs interact with EMT to promote glioma's development and progression are not fully understood, their elucidation is critical for developing effective diagnosis and treatments.

The transdifferentiation of epithelial cells into motile mesenchymal cells—a process known as epithelial-mesenchymal transition (EMT)—plays an important role in development, wound healing, and stem cell behavior, and also contributes to the pathology of fibrosis and cancer progression [7]. Tumor cells undergoing EMT acquire resistance to apoptosis and antitumor drugs while disseminating to distant sites [8]. Thus, EMT is a hallmark of carcinogenesis, and strategies targeting EMT pathways have therapeutic potential for cancer treatment [9].

Long noncoding (lnc)RNAs are RNA molecules longer than 200 bases that do not encode proteins [10; 11] and have been shown to function as master regulators of gene expression in normal biological processes and in diseases such as cancer [12]. As an important component of competing endogenous (ce)RNA networks, lncRNAs act as micro (mi)RNA sponges to regulate protein encoding and translation in cancer [13; 14]. For example, Upregulated TMPO antisense transcript 1 is a ceRNA that sponges miR-140-5p in gastric cancer cells, thereby alleviating inhibition of SRY-box transcription factor 4, an EMT regulator [15]. Although lncRNAs related to EMT have been identified [16], little is known about their role in glioma. The lncRNA-microRNA-mRNA ceRNA networks link the function of protein-coding mRNAs with that of non-coding RNAs such as microRNA, long non-coding RNA, pseudogenic RNA, and circular RNA. We learned that LINC01133 inhibits gastric cancer progression and metastasis by acting as a ceRNA for miR-106a-3p to regulate adenomatous polyposis coli (APC) gene expression and the Wnt/ β -catenin pathway[17]. And the SMAD5-AS1 target miR-135b-5p to inhibits diffuse large B cell lymphoma proliferation in the same pathway[18]. So, the deregulation of ceRNA networks may an important factor in lead to human diseases including cancer[13].

In this study, we used data from The Cancer Genome Atlas (TCGA) to establish an EMT-related lncRNA signature for glioma that was subsequently validated using data from the Chinese Glioma Genome Atlas (CGGA). We also established a lncRNA-mediated ceRNA network to investigate the downstream targets and mechanisms of EMT-related lncRNAs (Fig. 1).

Material And Methods

Patients and datasets

Clinical information of glioma patients was obtained from TCGA (<https://cancergenome.nih.gov/>) and CGGA (<http://www.cgga.org.cn>) websites; the 2 datasets were used for training and validation, respectively. Clinicopathologic information of the study population is shown in **Fig. S1**.

LncRNA profiling

The lncRNA profile was determined using an established mining method[19]. Briefly, genes were identified as protein-coding or noncoding based on their Refseq IDs, and only long noncoding genes in the NetAffx annotation files were retained. A total of 14,142 lncRNAs were obtained from the TCGA dataset, among which 215 were EMT-related as determined by gene set enrichment analysis (GSEA) (<http://www.broadinstitute.org/gsea/index.jsp>) [20]. Ultimately, 162 EMT-related lncRNAs from TCGA and 152 EMT-related lncRNAs from CGGA were retained to construct EMT lncRNA coexpression networks ($P \leq 0.001$) using Cytoscape v3.7.0 software (The Cytoscape Consortium, San Diego, CA, USA).

Establishment of a lncRNA signature

To establish a lncRNA signature for predicting survival, univariate Cox analysis was carried out with the 2 datasets. We screened lncRNAs with prognostic significance and identified 16 EMT-related lncRNAs that were closely associated with glioma patient survival ($P < 0.01$). We then assigned a risk score using the

least absolute shrinkage and selection operator (LASSO) Cox regression algorithm [21; 22; 23] and defined the inclusion criteria for the 9 lncRNAs and their constants, selecting the optimal penalty parameter β for a minimum 10-fold cross validation in the training set. The risk score was calculated with the following formula:

$$\text{Risk score} = \beta_{\text{gene1}} \times \text{expr}_{\text{gene1}} + \beta_{\text{gene2}} \times \text{expr}_{\text{gene2}} + \dots + \beta_{\text{genen}} \times \text{expr}_{\text{genen}}$$

where $\text{expr}_{\text{genen}}$ is the expression level of a lncRNA. Low- and high-risk groups were distinguished based on median risk score.

Construction of a ceRNA network

The interaction of lncRNAs, miRNAs, and mRNAs in glioma was predicted based on overlapping miRNA seed sequence binding sites in lncRNAs and mRNAs. The interaction between the 9 identified EMT-related lncRNAs and miRNAs was first predicted using the MiRcode database, which provides whole-transcriptome human miRNA target predictions based on the GENCODE annotation and includes > 10,000 long noncoding RNA genes. miRTarBase, TargetScan, and miRDB databases were then used to identify aberrantly expressed miRNA–mRNA pairs. We analyzed only target mRNAs that matched the databases. Finally, lncRNA–miRNA and miRNA–mRNA pairs were merged based on shared miRNAs into a ceRNA network of EMT-related lncRNAs, which was visualized using Cytoscape v3.7.0 software.

Statistical analysis

The expression of EMT-related lncRNAs between the two risk groups and their relationship with WHO grades were analyzed using one-way ANOVA. To compare glioma characteristics with different clinical pathology risk scores, a one-way ANOVA or t-test was conducted to compare risk scores for patients grouped by clinical or molecular pathology characteristics. Uni- and multivariate Cox regression analyses and principal component analysis (PCA) were performed using R v3.6.3 and SPSS v22 software applications (SPSS Inc, Chicago, IL, USA). GSEA was performed to investigate the functions of DEGs between the 2 groups of gliomas. Survival status was determined by univariate Cox regression analysis; the Kaplan–Meier curve was generated using Prism v6 software (GraphPad, La Jolla, CA, USA). Operating characteristic (ROC) curves were used to study the prediction efficiency of the risk signature, age, grade IDH-mutant status, and 1p/19q codeletion status. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses were performed using the “clusterProfiler” package of R software ($P_{\text{adj}} < 0.05$ and $Q < 0.05$). GSEA was used for functional annotation of genes. A 2-sided P value < 0.05 was considered significant for all statistical tests.

Results

Identification of prognostic EMT-associated lncRNAs in glioma

We extracted 14,142 and 4304 lncRNAs from the TCGA and CGGA datasets, respectively, and determined their expression profiles. A total of 215 EMT-related genes were extracted by GSEA (HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION M5930, ANASTASSIOU_MULTICANCER_INVASIVENESS_SIGNATURE M2572). We identified 69 EMT-related lncRNAs by constructing EMT-related lncRNA coexpression networks ($\text{cor} = 0.6$, $P < 0.0001$). Ultimately, 162 and 152 EMT-related lncRNAs were identified from TCGA and CGGA, respectively.

Establishment of an EMT-related lncRNA signature in glioma

To identify prognostic lncRNAs, expression data for each lncRNA in TCGA and CGGA were analyzed with a univariate Cox proportional hazards regression model ($P < 0.01$). We selected 137 and 92 lncRNAs related to glioma patient prognosis from TCGA and CGGA, respectively. We screened lncRNAs with prognostic significance in both datasets and identified 16 that were significantly associated with overall survival (OS) of glioblastoma patients ($P < 0.01$). LASSO Cox regression was used to identify the optimal prognostic lncRNAs in the TCGA data set and establish the risk signature. Nine of the 16 candidate lncRNAs retained their prognostic significance and were thus selected as independent prognostic markers (Fig. 2a–c). All the 9 EMT-related lncRNAs, LINC00900, MIR210HG, MIR22HG, PVT1, and SNHG18 were found to be risky lncRNAs with $\text{HR} > 1$, and HAR1A, LINC00641, SLC25A21-AS1, and SNAI3-AS1 were protective lncRNAs ($\text{HR} < 1$) (Table 1). Based on the median risk score, we divided glioma patients into low- and high-risk groups (Fig. S2); OS was longer in the former than in the latter ($P < 0.0001$; Fig. 2d). Then, we use the CGGA($n = 508$) dataset as a validation dataset to calculate the risk scores. We found a significant difference in OS between the two risk groups ($P < 0.0001$; Fig. 2e).

Table 1
The 9 EMT-related lncRNAs identified from the lasso cox regression in TCGA dataset

Symbol	HR	Low95	High95	P value	β Value
HAR1A	0.26	0.20	0.36	5.49E-18	-0.015
LINC00641	0.37	0.32	0.43	7.35E-38	-0.053
LINC00900	6.35	4.70	8.59	2.67E-33	0.417
MIR210HG	2.20	1.86	2.61	7.51E-20	0.036
MIR22HG	2.33	1.96	2.77	5.13E-22	0.007
PVT1	4.87	3.82	6.20	9.08E-38	0.146
SLC25A21-AS1	0.25	0.20	0.31	8.65E-39	-0.410
SNAI3-AS1	0.12	0.07	0.21	1.54E-15	-0.366
SNHG18	2.66	2.30	3.06	1.79E-41	0.012

Validation of the 9-lncRNA prognostic risk signature

We ranked the risk scores of patients in the TCGA training set and analyzed their distribution and plotted survival status (Fig. 3a). The heatmap revealed differences in the expression patterns of the prognostic lncRNAs between low- and high-risk patient groups. The expression of 4 protective lncRNAs was upregulated whereas that of 5 risk-associated lncRNAs was downregulated in patients with low risk scores; the opposite trends were observed in patients with high risk scores. These results were confirmed in the CGGA validation set (Fig. 3b).

Expression of 9 EMT-related lncRNAs and relationship with patient prognosis

We evaluated the expression of the 9 EMT-related lncRNAs according to tumor grade in the TCGA training set and CGGA validation dataset. LINC00900, MIR210HG, MIR22HG, PVT1, and SNHG18 increased whereas HAR1A, LINC00641 SLC25A21-AS1, and SNAI3-AS1 decreased with tumor grade (Fig. 4a, c). LINC00900, MIR210HG, MIR22HG, PVT1, and SNHG18 levels were higher whereas those of HAR1A, LINC00641 SLC25A21-AS1, and SNAI3-AS1 were lower in the high-risk group than in the low-risk group (Fig. 4b, d). The results from the two data are consistent. We also examined the significance of these 9 lncRNAs for patient survival using the TCGA dataset. Based on the expression level of each lncRNA, the survival time associated with each gene differed significantly between high- and low-risk groups (Fig. S3). Uni- and multivariate Cox regression analyses showed that the 9-lncRNA signature was an independent factor significantly associated with OS (TCGA: $P = 0.041$, hazard ratio [HR] = 1.806; CGGA: $P < 0.001$, HR = 1.855) (Fig. 5c-f).

Prognostic performance of risk scores and relationship with clinical status of glioma patients

Receiver operating characteristic (ROC) curves of the TCGA dataset showed that risk score (area under the ROC curve [AUC] = 0.771), age (AUC = 0.823), tumor grade (AUC = 0.813), 1p/19q codeletion status (AUC = 0.638), and isocitrate dehydrogenase (*IDH*) gene mutation status (AUC = 0.847) could accurately predict survival in glioma patients (Fig. 5a). These results were confirmed using the CGGA dataset (risk score, AUC = 0.730; age, AUC = 0.630; tumor grade, AUC = 0.775; 1p/19q codeletion status, AUC = 0.613; and *IDH* mutation status, AUC = 0.726) (Fig. 5b), demonstrating that the prognostic signature was reliable and effective. Uni- and multivariate Cox regression analyses performed using the TCGA dataset showed that risk score, age, WHO grade, and *IDH* mutation status were all related to OS (Fig. 5c, d); this was confirmed using the CGGA validation dataset (Fig. 5e, f). These results suggest that risk score derived from EMT-related lncRNAs is an independent factor that predicts the prognosis of glioma patients.

PCA and functional annotation of EMT-related lncRNAs

We carried out PCA to investigate differences between low- and high-risk groups based on EMT-related and overall gene expression profiles (Fig. 6a–d). Low- and high-risk glioma patient groups were distinguished based on EMT gene expression levels. We also found that EMT status was associated with a specific lncRNA signature. We performed GO and KEGG pathway analyses of the top 3000 genes that

were differentially expressed between low- and high-risk groups and found that differentially expressed genes (DEGs) were enriched in the following GO terms: M phase of cell cycle, Apoptosis, Tumor necrosis factor-mediated signaling pathway, and Regulation of cell adhesion. Additionally, KEGG pathway analysis revealed that Focal adhesion, Adherens junction, Pathways in cancer, and Cell adhesion molecules were significantly enriched (Fig. 6e, f). Functional annotation by GSEA showed that DEGs between the 2 groups were enriched in Epithelial-mesenchymal transition, E2F targets, P53 Pathway and IL6/JAK/STAT3 signaling (Fig. 6g). These results indicate that the two risk groups strongly correlated with the malignancy of glioma cells.

LncRNA-mediated ceRNA network and mechanisms of glioma tumorigenesis

To further investigate the regulatory role of lncRNAs in glioma, a lncRNA–miRNA–mRNA network was established by integrating expression profile data and regulatory relationships. We identified 220 EMT-related lncRNA–miRNA pairs comprising 4 lncRNAs and 28 miRNAs. miRNA–mRNA pairs were predicted using the miRTarBase, TargetScan, and miRDB databases; all target mRNAs with matches in these databases were included in the analysis. In total, 145 target mRNAs (Table 2) were included in the coexpression network that combined lncRNA–miRNA and miRNA–mRNA pairs (Fig. 7a). This ceRNA network revealed complex lncRNA–miRNA–mRNA interactions that could potentially contribute to EMT in glioma. GO and KEGG pathway enrichment analyses showed that target genes of ceRNAs were mainly enriched in the following biological functions and pathways: Vasculogenesis, Adherens junction organization, Ras protein signal transduction, Mitogen-activated protein kinase signaling, miRNAs in cancer, Transcriptional misregulation in cancer, Focal adhesion, and Glioma (Fig. 7b, c). We also identified genes of related functions and pathways and examined the enrichment of each functional pathway (Fig. 7d–g). Thus, the EMT-related ceRNA network plays an important role in glioma progression.

Table 2
The EMT-related lncRNAs, miRNAs, and mRNAs included in the ceRNA network

The type of RNAs	Gene symbols
LncRNA	MIR22HG, HAR1A, MIR210HG, PVT1
miRNA	hsa-miR-23b-3p, hsa-miR-20b-5p, hsa-miR-363-3p, hsa-miR-17-5p hsa-miR-139-5p, hsa-miR-125b-5p, hsa-miR-27a-3p, hsa-miR-125a-5p hsa-miR-24-3p, hsa-miR-206, hsa-miR-140-5p, hsa-miR-129-5p, hsa-miR-107, hsa-miR-33a-3p, hsa-miR-1244, hsa-miR-10a-5p hsa-miR-216b-5p, hsa-miR-876-3p, hsa-miR-490-3p, hsa-miR-217 hsa-miR-425-5p, hsa-miR-135a-5p, hsa-miR-22-3p, hsa-miR-3619-5p hsa-miR-761, hsa-miR-212-3p, hsa-miR-375, hsa-miR-613
mRNA	TNFAIP3, KIF23, SLC12A5, LCOR, E2F2, MET, RHOC, PHLPP2, TRIP10, PIP4K2A, GINS4, RRM2, MMP11, KPNA2, ZNF217, SGMS1, SAMD12, SFRP1, PDGFRA, CDKN1A, RBM47, ELAVL2, RUNX3, PRKCE, BTG3 DLGAP5, SLC16A9, PAX3, SAT1, BCL7A, GOLGA8A, RAP1B, LAMC1, WDR37, KLF10, EIF4EBP1, HOXA3, VEGFA, CMTM4, LHFPL4, RAPGEF4, NACC2, FZD6, RPS6KA5, KLF12, AKAP11, PPIA, AMOT, MAP3K9, E2F1, KIAA0513, AMOTL2, HOXA1, TEF, CNN3, RNF165, SPRY4, SNCG, RIMS3, EZH2, HAS2, PTGFRN, NR3C2, SMOC1, SCAMP5, WEE1, DUSP5, TGFR3, CADM2, GNG12, DUSP10, TRPM6, IGF1R, NF1, HOXC6, CBX6, ELOVL2, GABBR2, MYT1, RPS6KA1, CPEB3, SPRY2, ERBB3, TRIM29, CALU, ARC, INMT, CCDC65, POLQ, SOD2, EXPH5, WT1, NR2F2, NET1, CA2, CRY2, ITGA2, KANK4, COL4A4, MMP2, FAM129A, PDIA5, LRRC2, MAP3K5, ENDOU, PDPN, CSRP3, ZNF25, MXI1, FAM84B, TPPP, PPP1R3B, FGF9, DDN, GATA2, EN2, MYO1B, ARPC5, F3, BCL11B, FBLIM1, BCR, KIAA1147, VDR, NR5A2, BRWD3, DENND5B, GABRB1, TAOK1, C1S, CNNM2, FGFR1, RUNX1T1, E2F7, SMAD7, ELAVL4, NR1D2, DACH1, ENPEP, KIAA1109, HMGB2, HOXA10, LATS2, PDE10A, RAP2A

Discussion

In this study, we identified a 9-lncRNA signature that could distinguish between low- and high-risk glioma patients based on the median risk score in TCGA. These 9 EMT-related lncRNAs have prognostic significance in glioma: patients in the low-risk group had longer OS than those in the high-risk group. Deletion of chromosomal arms 1p and/or 19q and *IDH* mutations are considered as useful biomarkers in glioma [2; 24; 25]; we evaluated the risk scores of *IDH* and 1p/19q codeletion in the TCGA dataset and found that both were related to OS. The results were confirmed using the CGGA dataset, indicating that the 9-lncRNA signature is reliable and effective for predicting prognosis. We also established a lncRNA–miRNA–mRNA ceRNA network to predict the regulatory relationships of the 9 lncRNAs; functional enrichment analysis of downstream target genes revealed that the lncRNAs are involved in tumorigenesis and cancer progression.

As an intricate genetic procedure, EMT can be utilized by both normal and tumor epithelial cells to enable them to separate from neighboring cells and migrate[26], and have influenced invasive, cancer progression and therapy resistance property of cancer cells[27; 28]. Previous reports indicate that Snail[29; 30], ZEB[31; 32], Twist[33] families and some other signaling pathways genes[34; 35; 36] have been identified as master regulators in EMT progress. More and more research shows lncRNAs epigenetically orchestrate EMT, and have promote our understanding of the molecular mechanisms that lncRNA-mediated in cancer initiation and malignancy progression[6]. lncRNAs play a critical role in diverse cellular processes in both normal and disease states [12; 37]. Specifically, lncRNAs promote or act as decoys to repress transcription, or function as epigenetic regulators or scaffolds that interact with various proteins to form ribonucleoprotein complexes [38; 39]. lncRNAs are involved in multiple signaling pathways including p53, nuclear factor κ B (NF- κ B), phosphatidylinositol 3-kinase/protein kinase B, and Notch, which have been linked to EMT. Among the 9 EMT-related lncRNAs identified in this study, LINC00900, MIR210HG, MIR22HG, PVT1, and SNHG18 were associated with increased risk of glioma whereas HAR1A, LINC00641, SLC25A21-AS1, and SNAI3-AS1 were protective. PVT1 is upregulated in various carcinomas, and its overexpression is associated with poor survival in patients [40]. PVT1 functions act as a ceRNA to protect mRNAs from miRNA-mediated repression and associated with the development of resistance to common chemotherapeutic agents[41; 42]. MIR210HG sponges miR-1226-3p to facilitate breast cancer cell invasion and metastasis and EMT via regulation of mucin-1c; and MIR210HG upregulation was shown to be associated with advanced International Federation of Gynecology and Obstetrics (FIGO) stage, metastasis, and poor prognosis in cervical cancer [43; 44]. SNHG18 promotes resistance of glioma to radiotherapy by repressing semaphorin 5A [45]. LINC00641 expression level was found to be decreased in breast cancer tissue, which was negatively related to tumor size, lymph node metastasis, and clinical stage [46]. SNAI3-AS1 could bond up-frameshift protein 1 (UPF1), regulate Smad7 expression, and activate TGF- β /Smad pathway, to induce EMT in hepatocellular carcinoma[47], and it also acts an important factor in the ceRNA network in Adipose Tissue From Type 2 Diabetes[48]. Therefore, the close relationship between the lncRNA and EMT we screened and the important role of lncRNA in the ceRNA network have been further confirmed in other tumors or diseases.

Although some of the identified lncRNAs have not been previously reported as associated with EMT, our results suggest that they may directly or indirectly be involved in this process. There were some limitations to our study. Firstly, our analyses were based on publically available datasets and require confirmation in other patient populations. Moreover, experiments should be carried out to validate the roles of the identified lncRNAs and their targets and the lncRNA–miRNA–mRNA ceRNA network in glioma.

Conclusions

The 9-EMT–related lncRNA signature established in this study has prognostic value for glioma patients, while the ceRNA network provides insight into the molecular basis of glioma progression as well as potential therapeutic targets for its treatment.

Abbreviations

TCGA

The Cancer Genome Atlas; CGGA:The Chinese Glioma Genome Atlas; EMT:epithelial-mesenchymal transition; OS:overall survival; GO:gene ontology; MF:Molecular function; KEGG:Kyoto Encyclopedia of Genes and Genomes; AUC:Area under the curve; GSEA:Gene set enrichment analysis ; ROC:Receiver operating characteristic curve;

Declarations

Acknowledgments

The authors gratefully acknowledge contributions from the CGGA network and the TCGA Network.

Funding

The present study was supported by the National Natural Science Foundation (grant nos. 81660420, 81960456, and 81960457), the Natural Science Foundation of Jiangxi Province (grant no. 20171ACB20035) and the Jiangxi Province Department of Education Science and technology research project (Grant/Award number: GJJ190018).

Availability of data and materials

The datasets analysed during the current study are available in the TCGA (<http://cancergenome.nih.gov/>) and CGGA (<http://www.cgga.org.cn/>) repository.

Author contributions

XZ and KH designed the research. CT, QH, and HL created tables, and figures, and were involved in manuscript writing and proofreading. HG gave analysis advice, and performed the literature search. XZ and KH supervised the research and critically read the draft manuscript. All authors proofread and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Jovcevska I, Kocevar N, Komel R. Glioma and glioblastoma - how much do we (not) know? *Mol Clin Oncol*. 2013;1:935–41.
2. Louis DN, Perry A, Reifenberger G, von Deimling A, Figarella-Branger D, Cavenee WK, Ohgaki H, Wiestler OD, Kleihues P, Ellison DW. The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. *Acta Neuropathol*. 2016;131:803–20.
3. Fedele M, Cerchia L, Pegoraro S, Sgarra R, Manfioletti G. Proneural-Mesenchymal Transition: Phenotypic Plasticity to Acquire Multitherapy Resistance in Glioblastoma. *Int J Mol Sci* 20 (2019).
4. Iser IC, Pereira MB, Lenz G, Wink MR. The Epithelial-to-Mesenchymal Transition-Like Process in Glioblastoma: An Updated Systematic Review and In Silico Investigation. *Med Res Rev*. 2017;37:271–313.
5. Kaufhold S, Bonavida B. Central role of Snail1 in the regulation of EMT and resistance in cancer: a target for therapeutic intervention. *J Exp Clin Cancer Res*. 2014;33:62.
6. Cheng JT, Wang L, Wang H, Tang FR, Cai WQ, Sethi G, Xin HW. and Z. Ma, Insights into Biological Role of LncRNAs in Epithelial-Mesenchymal Transition. *Cells* 8 (2019).
7. Lamouille S, Xu J, Derynck R. Molecular mechanisms of epithelial-mesenchymal transition. *Nat Rev Mol Cell Biol*. 2014;15:178–96.
8. Zhang Z, Wang S, Liu W. EMT-related long non-coding RNA in hepatocellular carcinoma: A study with TCGA database. *Biochem Biophys Res Commun*. 2018;503:1530–6.
9. Mittal V. Epithelial Mesenchymal Transition in Tumor Metastasis. *Annu Rev Pathol*. 2018;13:395–412.
10. Kopp F, Mendell JT. Functional Classification and Experimental Dissection of Long Noncoding RNAs. *Cell*. 2018;172:393–407.
11. Huang Y. The novel regulatory role of lncRNA-miRNA-mRNA axis in cardiovascular diseases. *J Cell Mol Med*. 2018;22:5768–75.
12. Peng WX, Koirala P, Mo YY. LncRNA-mediated regulation of cell signaling in cancer. *Oncogene*. 2017;36:5661–7.

13. Qi X, Zhang DH, Wu N, Xiao JH, Wang X, Ma W. ceRNA in cancer: possible functions and clinical implications. *J Med Genet.* 2015;52:710–8.
14. Tay Y, Rinn J, Pandolfi PP. The multilayered complexity of ceRNA crosstalk and competition. *Nature.* 2014;505:344–52.
15. Sun Y, Han C, Non-Coding L. RNA TMPO-AS1 Promotes Cell Migration and Invasion by Sponging miR-140-5p and Inducing SOX4-Mediated EMT in Gastric Cancer. *Cancer Manag Res.* 2020;12:1261–8.
16. Mitra R, Chen X, Greenawalt EJ, Maulik U, Jiang W, Zhao Z, Eischen CM. Decoding critical long non-coding RNA in ovarian cancer epithelial-to-mesenchymal transition. *Nat Commun.* 2017;8:1604.
17. Yang XZ, Cheng TT, He QJ, Lei ZY, Chi J, Tang Z, Liao QX, Zhang H, Zeng LS, Cui SZ. LINC01133 as ceRNA inhibits gastric cancer progression by sponging miR-106a-3p to regulate APC expression and the Wnt/beta-catenin pathway. *Mol Cancer.* 2018;17:126.
18. Zhao CC, Jiao Y, Zhang YY, Ning J, Zhang YR, Xu J, Wei W, Kang-Sheng G. Lnc SMAD5-AS1 as ceRNA inhibit proliferation of diffuse large B cell lymphoma via Wnt/beta-catenin pathway by sponging miR-135b-5p to elevate expression of APC. *Cell Death Dis.* 2019;10:252.
19. Zhang X, Sun S, Pu JK, Tsang AC, Lee D, Man VO, Lui WM, Wong ST, Leung GK. Long non-coding RNA expression profiles predict clinical phenotypes in glioma. *Neurobiol Dis.* 2012;48:1–8.
20. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A.* 2005;102:15545–50.
21. Bovelstad HM, Nygard S, Storvold HL, Aldrin M, Borgan O, Frigessi A, Lingjaerde OC. Predicting survival from microarray data—a comparative study. *Bioinformatics.* 2007;23:2080–7.
22. Sauerbrei W, Royston P, Binder H. Selection of important variables and determination of functional form for continuous predictors in multivariable model building. *Stat Med.* 2007;26:5512–28.
23. Hu X, Martinez-Ledesma E, Zheng S, Kim H, Barthel F, Jiang T, Hess KR, Verhaak RGW. Multigene signature for predicting prognosis of patients with 1p19q co-deletion diffuse glioma. *Neuro Oncol.* 2017;19:786–95.
24. Wirsching HG, Weller M. The Role of Molecular Diagnostics in the Management of Patients with Gliomas. *Curr Treat Options Oncol.* 2016;17:51.
25. Bai H, Harmanci AS, Erson-Omay EZ, Li J, Coskun S, Simon M, Krischek B, Ozduman K, Omay SB, Sorensen EA, Turcan S, Bakirciglu M, Carrion-Grant G, Murray PB, Clark VE, Ercan-Sencicek AG, Knight J, Sencar L, Altinok S, Kaulen LD, Gulez B, Timmer M, Schramm J, Mishra-Gorur K, Henegariu O, Moliterno J, Louvi A, Chan TA, Tannheimer SL, Pamir MN, Vortmeyer AO, Bilguvar K, Yasuno K, Gunel M. Integrated genomic characterization of IDH1-mutant glioma malignant progression. *Nat Genet.* 2016;48:59–66.
26. Nieto MA, Huang RY, Jackson RA, Thiery JP. EMT: 2016 Cell. 2016;166:21–45.
27. Kalluri R, Weinberg RA. The basics of epithelial-mesenchymal transition. *J Clin Investig.* 2009;119:1420–8.

28. Cebria-Costa JP, Millanes-Romero A, de Herreros AG, Peiro S. The Epithelial-to-Mesenchymal Transition (EMT), a Particular Case. *Mol Cell Oncol*. 2014;1:e960770.
29. Battistelli C, Cicchini C, Santangelo L, Tramontano A, Grassi L, Gonzalez FJ, de Nonno V, Grassi G, Amicone L, Tripodi M. The Snail repressor recruits EZH2 to specific genomic sites through the enrollment of the lncRNA HOTAIR in epithelial-to-mesenchymal transition. *Oncogene*. 2017;36:942–55.
30. Myung JK, Choi SA, Kim SK, Wang KC, Park SH. Snail plays an oncogenic role in glioblastoma by promoting epithelial mesenchymal transition. *Int J Clin Exp Pathol*. 2014;7:1977–87.
31. Siebzehnruhl FA, Silver DJ, Tugertimur B, Deleyrolle LP, Siebzehnruhl D, Sarkisian MR, Devers KG, Yachnis AT, Kupper MD, Neal D, Nabils NH, Kladde MP, Suslov O, Brabletz S, Brabletz T, Reynolds BA, Steindler DA. The ZEB1 pathway links glioblastoma initiation, invasion and chemoresistance. *EMBO Mol Med*. 2013;5:1196–212.
32. Qi S, Song Y, Peng Y, Wang H, Long H, Yu X, Li Z, Fang L, Wu A, Luo W, Zhen Y, Zhou Y, Chen Y, Mai C, Liu Z, Fang W. ZEB2 mediates multiple pathways regulating cell proliferation, migration, invasion, and apoptosis in glioma. *PLoS One*. 2012;7:e38842.
33. Mikheeva SA, Mikheev AM, Petit A, Beyer R, Oxford RG, Khorasani L, Maxwell JP, Glackin CA, Wakimoto H, Gonzalez-Herrero I, Sanchez-Garcia I, Silber JR, Horner PJ, Rostomily RC. TWIST1 promotes invasion through mesenchymal change in human glioblastoma. *Mol Cancer*. 2010;9:194.
34. Derynck R, Muthusamy BP, Saeteurn KY. Signaling pathway cooperation in TGF-beta-induced epithelial-mesenchymal transition. *Curr Opin Cell Biol*. 2014;31:56–66.
35. Guo F, Wang Y, Liu J, Mok SC, Xue F, Zhang W. CXCL12/CXCR4: a symbiotic bridge linking cancer cells and their stromal neighbors in oncogenic communication networks. *Oncogene*. 2016;35:816–26.
36. Kahlert UD, Maciarczyk D, Doostkam S, Orr BA, Simons B, Bogiel T, Reithmeier T, Prinz M, Schubert J, Niedermann G, Brabletz T, Eberhart CG, Nikkhah G, Maciarczyk J. Activation of canonical WNT/beta-catenin signaling enhances in vitro motility of glioblastoma cells by activation of ZEB1 and other activators of epithelial-to-mesenchymal transition. *Cancer letters*. 2012;325:42–53.
37. Qian X, Zhao J, Yeung PY, Zhang QC, Kwok CK. Revealing lncRNA Structures and Interactions by Sequencing-Based Approaches. *Trends Biochem Sci*. 2019;44:33–52.
38. Mercer TR, Dinger ME, Mattick JS. Long non-coding RNAs: insights into functions. *Nat Rev Genet*. 2009;10:155–9.
39. Ulitsky I, Bartel DP. lincRNAs: genomics, evolution, and mechanisms. *Cell*. 2013;154:26–46.
40. Tseng YY, Moriarity BS, Gong W, Akiyama R, Tiwari A, Kawakami H, Ronning P, Reuland B, Guenther K, Beadnell TC, Essig J, Otto GM, O'Sullivan MG, Largaespada DA, Schwertfeger KL, Marahrens Y, Kawakami Y, Bagchi A. PVT1 dependence in cancer with MYC copy-number increase. *Nature*. 2014;512:82–6.
41. Xiao M, Feng Y, Liu C, Zhang Z. Prognostic values of long noncoding RNA PVT1 in various carcinomas: An updated systematic review and meta-analysis. *Cell Prolif*. 2018;51:e12519.

42. Ogunwobi OO, Kumar A. Chemoresistance Mediated by ceRNA Networks Associated With the PVT1 Locus. *Frontiers in oncology*. 2019;9:834.
43. Wang AH, Jin CH, Cui GY, Li HY, Wang Y, Yu JJ, Wang RF, Tian XY. MIR210HG promotes cell proliferation and invasion by regulating miR-503-5p/TRAF4 axis in cervical cancer. *Aging*. 2020;12:3205–17.
44. Li XY, Zhou LY, Luo H, Zhu Q, Zuo L, Liu GY, Feng C, Zhao JY, Zhang YY, Li X. The long noncoding RNA MIR210HG promotes tumor metastasis by acting as a ceRNA of miR-1226-3p to regulate mucin-1c expression in invasive breast cancer. *Aging*. 2019;11:5646–65.
45. Zheng R, Yao Q, Ren C, Liu Y, Yang H, Xie G, Du S, Yang K, Yuan Y. Upregulation of Long Noncoding RNA Small Nucleolar RNA Host Gene 18 Promotes Radioresistance of Glioma by Repressing Semaphorin 5A. *Int J Radiat Oncol Biol Phys*. 2016;96:877–87.
46. Mao Q, Lv M, Li L, Sun Y, Liu S, Shen Y, Liu Z, Luo S. Long intergenic noncoding RNA 00641 inhibits breast cancer cell proliferation, migration, and invasion by sponging miR-194-5p. *J Cell Physiol*. 2020;235:2668–75.
47. Li Y, Guo D, Ren M, Zhao Y, Wang X, Chen Y, Liu Y, Lu G, He S. Long non-coding RNA SNAI3-AS1 promotes the proliferation and metastasis of hepatocellular carcinoma by regulating the UPF1/Smad7 signalling pathway. *J Cell Mol Med*. 2019;23:6271–82.
48. Hu W, Ding Y, Wang S, Xu L, Yu H, The Construction and Analysis of the Aberrant lncRNA-miRNA-mRNA Network in Adipose Tissue from Type 2 Diabetes Individuals with Obesity. *J Diabetes Res* 2020 (2020) 3980742.

Figures

Figure 1

Study flowchart.

Figure 2

Risk signature consisting of 9 EMT-related lncRNAs. (a, b) Inclusion criteria for the 9 EMT-related lncRNAs and their constants; the optimal penalty parameter β was selected for a minimum 10-fold cross validation of the training set. (c) Seven EMT-related lncRNAs. (d, e) Kaplan–Meier OS curves of patients in the TCGA (d) and CGGA (e) datasets assigned to high- and low-risk glioma groups based on risk score. (f) Heatmap

of the 9 EMT-related lncRNAs in low- and high-risk glioma. The distribution of clinicopathologic features was compared between low- and high-risk groups. **P<0.01, ***P<0.001.

Figure 3

Distribution of risk score, OS, and gene expression in the TCGA and CGGA datasets. (a, b) Distribution of risk score and OS. (c) Heat map of the expression of 9 EMT-related lncRNAs in low- and high-risk groups. Rows and columns show lncRNAs and patients, respectively.

Figure 4

Expression of 9 EMT-related lncRNAs according to tumor grade and risk group. ***P<0.001.

Figure 5

Predictive efficiency of the 9 EMT-related lncRNA glioma risk signature. (a, b) Association between WHO grade, age, 1p19q codeletion status, IDH mutation status, and survival rate. (c–f) Uni- and multivariate Cox regression analyses of the association between clinicopathologic factors (including risk score) and OS of patients in the TCGA (c, d) and CGGA (e, f) datasets.

Figure 6

Distinct EMT status in low-risk vs high-risk glioma patients. (a) PCA between low- and high-risk glioma patient groups based on total gene expression profiles. (b) PCA between low- and high-risk groups based on EMT-related genes. (c, d) PCA of low- and high-risk groups based on all candidate EMT-related lncRNAs (c) and the 9 identified EMT-related lncRNAs (d). (e, f) GO and KEGG pathway enrichment analyses of DEGs in low- and high-risk groups. (g) Significant enrichment of EMT-associated functions in the high-risk group, as determined by GSEA.

Figure 7

EMT-related ceRNA network and functional analysis of EMT-related lncRNAs. (a) Global view of the ceRNA network in glioma. Circles, triangles, and diamonds represent lncRNAs, miRNAs, and mRNAs, respectively. (b, c) GO (b) and KEGG pathway enrichment (c) analyses of mRNAs in the network. The horizontal axis shows the number of enriched genes and the color intensity represents the corrected P

value. (d, f) Genes in significant bioprocesses. (e, g) Cluster analysis of genes and pathways with related functions, and enrichment of each functional pathway.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Fig.S2.tif](#)
- [Fig.S3.tif](#)
- [Fig.S1.tif](#)