LINGO1, C7orf31 and VEGFA are prognostic genes of primary glioblastoma: analysis of gene expression microarray

S. LIU1, Y. XU2,*, S. ZHANG3

1Department of Rehabilitation Medicine, Jinan Central Hospital Affiliated to Shandong University, Jinan, Shandong 250013, P.R. China; 2Reviewing office, Jinan Central Hospital Affiliated to Shandong University, Jinan, Shandong 250013, P.R. China; 3ENT, Jinan Central Hospital Affiliated to Shandong University, Jinan, Shandong 250013, P.R. China

*Correspondence: youjunyxuyjx@sina.com

Received July 22, 2017/ Accepted November 3, 2017

Glioblastoma is the most prevalent malignant glioma in WHO grade IV and its median overall survival is 12–15 months. This study identifies the primary glioblastoma. prognostic genes. Gene expression data in primary glioblastomas with short-term (<12 months, N=16), intermediate (12–36 months, N=31), and long-term (>36 months, N=23) overall survival were downloaded from Gene Expression Omnibus (GSE53733). Limma determined the differentially expressed genes (DEGs) between different groups (|log2 fold change| ≥0.5 and p-value <0.05). STRING database and Cytoscape were used to predict protein-protein interactions between DEGs and to construct the PPI network (PPI, medium confidence ≥0.4), and CytoNCA plugin in Cytoscape calculated each DEG's degree, betweenness, sub-graph and closeness centralities. Long-term/short-term survival-related DEGs were defined as those with increased/decreased expression values and survival time. The following DEGs were identified; 161 between intermediate and short-term glioblastomas, 465 between long-term and short-term and 624 between long-term and intermediate tumors. The common FLRT1 and LINGO1 up-regulated DEGs and common down-regulated C7orf31 were identified in these three DEG sets. PPI networks were established, and VEGFA was the key DEG in each PPI network. The short-term survival-related DEGs were enriched in 3 cancer-related pathways. Moreover, FLRT1 and LINGO1 were long-term survival-related DEGs and C7orf31 and VEGFA were short-term survival DEGs. LINGO1, C7orf31, and VEGFA were confirmed using a further dataset, and we therefore conclude that LINGO1 might be a positive primary glioblastoma prognostic gene and C7orf31 and VEGFA might be negative prognosticators.

Key words: differentially expressed gene, overall survival, primary glioblastoma, prognosis, protein–protein interaction

Glioblastoma multiforme is the most prevalent and aggressive form of malignant glioma in WHO grade IV. It accounts for 56.1% of glioma and 47.1% of primary malignant tumors in the brain with morbidity of 3.2/100,000 population [1]. Primary glioblastoma appears more often in the elderly at median 64 years of age [1]. Despite surgical, chemotherapy and radiotherapy advances, glioblastoma 5-year survival rate remains less than 10%, and median overall survival time is still 12–15 months [2]. A better understanding of survival-related factors will aid development of a more targeted and personalized therapy for this cancer.

Results of investigation into genetic factors associated with glioblastoma's prognosis and survival time include: (1) C228T (−124 bp upstream start codon), a promoter mutation in telomerase reverse transcriptase (TERT) appears in 64.5% (75%×86%) of primary glioblastoma and it is related to shorter overall survival time (C228T vs. wild: 11 vs. 20 months, p=0.002) [3] and (2) C250T (−146 bp) in TERT promoter appears in 21.5% (25%×86%) of primary glioblastoma and it also associates with shorter overall survival time (C250T vs. wild: 12 vs. 20 months, p=0.04) [3]. In glioblastoma patients with wild-type TERT, the overall survival in the epidermal growth factor receptor (EGFR)-wild group is significantly longer than in the EGFR-amplification group (26.6 vs. 13.3 months, p=0.005) [4]. In addition, TERT promoter mutation is associated with EGFR amplification, cyclin-dependent kinase inhibitor 2A (CDKN2A) deletion and isocitrate dehydrogenase enzyme (IDH) wild-type and chromosome 10q loss [4].

Meta-analysis of 24 studies revealed that methylation of the O-6-methylguanine-DNA methyltransferase (MGMT) promoter positively correlates with longer overall survival [hazard ratio (HR)=0.48, I²=79.78] and progression-free survival (HR=0.43, I²=50.38) [5]. Meanwhile, programmed
cell death ligand 1 (PD-L1) is expressed in >88% of glioblastomas and its high expression is significantly related with shorter survival (p=0.028) [6]. Independent and negative predictors of overall survival include PD-L1 (HR=1.52; p=0.0343) [6] and high glioblastoma expression of the zeste homology enhancer 2 (EZH2) [7]. E3 ubiquitin-protein ligase MDM2 overexpression is also found more frequently in short-term survivors (≤1.5 years) than in the long-term (≥3 years) [8].

Despite these advances, the majority of genetic factors for primary glioblastoma prognosis remain sporadic or undiscovered. A genome-wide expression analysis on primary glioblastomas with different survivals should help identify more prognostic factors. Although Reifenberger et al. defined the molecular characters of long-term survivors using transcriptome-wide profiling data [9], they mainly focused on glioblastoma patients with IDH1/2-mutant, and there were only 12 of these successfully analyzed in the 70 glioblastoma cohort. In addition, only clustering was performed and no prognostic gene was identified except IDH1/2 mutations and the MGMT promoter methylation gene [9]. Herein, we re-analyzed the gene expression data for both IDH1/2 wild-type and IDH1/2-mutant glioblastomas up-loaded by Reifenberger et al. and investigated their relationship to primary glioblastoma prognosis.

Materials and methods

Gene expression data. Human gene expression microarray data (accession ID: GSE53733) [9] in primary glioblastoma tissues was obtained from Gene Expression Omnibus (GEO). Samples included 16 glioblastomas with short-term overall survival (<12 months; 1 IDH1/2-mutant and 15 IDH1/2-wildtype), 31 glioblastomas with intermediate overall survival (12–36 months; 2 IDH1/2-mutant and 29 IDH1/2-wildtype), and 23 glioblastomas with long-term overall survival (>36 months; 9 IDH1/2-mutant and 14 IDH1/2-wildtype). Frozen tissues were generated from the initial operation prior to the use of radiotherapy or chemotherapy treatment. Only specimens with a histologically estimated tumor cell content ≥80% were used, and the corresponding Affymetrix Human Genome U133 Plus 2.0 Array platform was also obtained from GEO.

Raw-data pre-processing. The raw microarray data in GSE53733 [9] was read and preprocessed using R software package affy (version: 1.54.0; available at: http://www.bioconductor.org/packages/release/bioc/html/affy.html). This process included background correction, data normalization and calculation of probe expressions. The following were then used to transform probe values to gene-symbol values: R annotation packages hgu133plus2.db (version: 3.2.3; available at: http://www.bioconductor.org/packages/release/data/annotation/html/hgu133plus2.db.html), annotate (version: 1.54.0; available at: http://www.bioconductor.org/packages/release/bioc/html/annotate.html), and org.Hs.eg.db (version: 3.4.1; available at: http://www.bioconductor.org/packages/release/data/annotation/html/org.Hs.eg.db.html). For a certain gene-symbol with multiple probes, the average probe value was considered the gene-symbol value. Probes corresponding to no proper gene-symbol were deleted.

Differential expression analysis. The limma R software package (version: 3.32.2; available at: http://www.bioconductor.org/packages/release/bioc/html/limma.html) [10] identified differentially expressed genes (DEGs) in 31 intermediate glioblastomas and 16 short-term glioblastomas, 23 long-term glioblastomas and 16 short-term glioblastomas and also in 23 long-term glioblastomas and 31 intermediate glioblastomas. P-values were calculated by limma t-test and adjusted using the Bayesian method. Significant DEGs between the groups were defined as the genes with |log₂ fold change (FC)| ≥0.5 and p<0.05.

Function and pathway enrichment analyses. DAVID (version: 6.8; available at: https://david.ncifcrf.gov/) [11] identified the Gene Ontology (GO; available at: http://www.geneontology.org/) functions and Kyoto Encyclopedia of Genes and Genomes (KEGG, available at: http://www.kegg.jp/) pathways associated with DEGs. DEG count ≥2 and p<0.05 in hypergeometric test were set as criteria in these analyses.

Common DEGs. Venny (version 2.1.0; available at: http://bioinfogp.cnb.csic.es/tools/venny/) compared the 3 DEG sets and identified common DEGs.

Protein-protein interactions (PPIs) between DEGs. STRING database (available at: http://string-db.org/) predicted PPIs between DEGs in each set based on neighborhood information, co-expression experiments, gene fusion, text-mining and co-occurrence information in STRING [12]. Cytoscape Software (version: 3.5.0; available at: http://www.cytoscape.org/) visualized PPIs with medium confidence ≥0.4 and PPI network construction was instituted [13]. The degrees of centrality, betweenness centrality, subgraph centrality, and closeness centrality correlate with the importance of nodes in the network. CytoNCA plugin (version: 2.1.6; available at: http://apps.cytoscape.org/apps/cytonca) [14] in Cytoscape was then utilized to identify key PPI network nodes by calculating the 4 centrality scores of each DEG.

Survival-related DEGs. Expression values in the 16 short-term glioblastomas were averaged for each DEG in the 3 DEG sets, and expression values in the 31 intermediate glioblastomas and the 23 long-term glioblastomas were also averaged. Long-term survival-related DEGs were defined as those with increasing expression values with survival time, and short-term survival-related DEGs were those with decreasing expression values.

Validation of key genes. A further set of human gene expression data (GSE50161) was downloaded from GEO to validate key genes identified by GSE53733. Samples included 13 normal brain tissues and 34 primary glioblastoma tissues from the initial operations. GSE50161 and GSE53733 used
the same platform and methods to pre-process raw data and perform differential expression analysis.

**Results**

**DEGs and the 3 DEG sets.** An expression matrix with 20,514 gene-symbols was constructed following data pre-processing. Differential expression analysis (log FC ≥0.5 and p-value <0.05) identified 161 DEGs between intermediate and short-term glioblastomas (87 up-regulated DEGs and 74 down-regulated), 465 DEGs between long-term and short-term glioblastomas (180 up-regulated and 285 down-regulated) and 624 DEGs between long-term and intermediate glioblastomas (153 up-regulated and 471 down-regulated).

**Functions and pathways enriched by DEGs.** The following regulated DEGs were established:

(A) in the intermediate vs. short-term glioblastomas: (1) up-regulated DEGs were enriched in 21 Biological Process (BP) terms, 9 Cellular Component (CC) terms, 6 Molecular Function (MF) terms (Figure 1, DEG count ≥2 and p-value <0.05) and 5 KEGG pathways (Figure 3A) and (2) the down-regulated DEGs participated in 11 BP terms, 4 CC terms, 2 MF terms (Figure 2) and the non-KEGG pathways (Figure 3A);

(B) in long-term vs. short-term glioblastomas: (1) up-regulated DEGs enriched in 17 BP terms, 11 CC terms, 12 MF terms (Figure 1) and 2 KEGG pathways (Figure 3A) and (2) the down-regulated DEGs participated in 45 BP terms, 17 CC terms, 15 MF terms (Figure 2) and 5 KEGG pathways (Figure 3A);

(C) in long-term vs. intermediate glioblastomas: (1) up-regulated DEGs enriched in 23 BP terms, 11 CC terms, 7 MF terms (Figure 1) and 8 KEGG pathways (Figure 3A) and (2) down-regulated DEGs participated in 152 BP terms, 32 CC terms, 19 MF terms (Figure 2) and 25 KEGG pathways (Figure 3A).

Most importantly, the up-regulated DEGs in the 3 DEG sets were all enriched in “cell adhesion”, “plasma membrane”, “integral plasma membrane components” and “integral membrane components” (Figure 1), while down-regulated DEGs all participated in “extracellular matrix organization”, “extracellular matrix”, “proteinaceous extracellular matrix”, “extracellular space” and the “extracellular region” (Figure 2).

**Common DEGs.** Although differences existed in the 3 DEG sets, the 2 up-regulated FLRT1 and LINGO1 DEGs and the C7orf31 down-regulated DEG were commonly identified (Figure 3B).

**PPI networks.** PPI networks were constructed for the following:

(1) the 161 DEGs between intermediate and short-term glioblastomas (Figure 4A). This included 74 DEGs (nodes) and 76 PPIs (edges, medium confidence ≥0.4). In particular, the up-regulated EGFR, RET, and BCAN DEGs were included in the top 10 DEGs based on the 4 centrality scores, together with the down-regulated MDM2 and GHR DEGs (Figure 4B); (2) the 465 DEGs between long-term and short-term glioblastomas (Figure 5A). This involved 254 DEGs (nodes) and 376 PPIs (edges). Here, the down-regulated VEGFA and MEIS1 DEGs were involved in the top 10 based on the 4 centrality scores (Figure 5B); (A) the 624 DEGs between long-term and intermediate glioblastomas (Figure 6A). This had 436 DEGs (nodes) and 1300 PPIs (edges), and included the down-regulated TSPO, IL6, EGFR, VEGFA, FOs and EGR1 DEGs in the top 10 based on the 4 centrality scores (Figure 6B).

These were the key DEGs in the PPI networks.

**Survival-related DEGs.** Totals of 171 long-term survival-related DEGs and 237 short-term survival-related DEGs were identified by expression values in the short-term, intermediate, and long-term glioblastomas. The common up-regulated FLRT1 and LINGO1 are long-term survival-related DEGs and the common down-regulated C7orf31 and key PPI network VEGFA DEG were related to short-term survival.

The long and short-term survival-related DEGs were enriched in 3 and 5 pathways, respectively and the “Proteoglycans in cancer” (VEGFA), “Basal cell carcinoma”, and “Pathways in cancer” (VEGFA) examples of these pathways are highlighted in Figure 3C.

**Validation of key genes.** The above-mentioned analysis identified FLRT1, LINGO1, VEGFA and 27orf31 as key genes. While GSE50161 differential expression analysis determined that FLRT1 was not differentially expressed in normal brain and glioblastoma tissues, LINGO1 was significantly down-regulated in glioblastoma, and this accords with LINGO1 being a long-term survival-related DEG. In addition, VEGFA and C7orf31 were significantly up-regulated in glioblastoma, thus supporting their status as DEGs related to short-term survival.

**Discussion**

This study analyzed gene expression data in primary glioblastomas with short-term, intermediate and long-term survival rates. Different DEGs, functions and pathways were distinguished in different glioblastomas with FLRT1 and LINGO1 usually up-regulated and C7orf31 normally down-regulated.

PPI networks were then constructed with the following results: (1) EGFR, RET, BCAN, MDM2 and GHR were key DEGs between intermediate and short-term glioblastomas; (2) VEGFA and MEIS1 were key DEGs between long-term and short-term glioblastomas and (3) TSPo, IL6, EGFR, VEGFA, FOs and EGR1 were key DEGs between long-term and intermediate glioblastomas. In addition, FLRT1, LINGO1, C7orf31 and VEGFA were survival-related DEGs.
Figure 1. Gene Ontology functions enriched by up-regulated DEGs. DEGs: differentially expressed genes; BP: Biological Process; CC: Cellular Component; MF: Molecular Function.
Figure 2. Gene Ontology functions enriched by down-regulated DEGs: Biological Process (BP), Cellular Component (CC), Molecular Function (MF).
and LINGO1, C7orf31 and VEGFA were further confirmed with another dataset of normal brain and glioblastoma tissues.

The relationships between MDM2/EGFR and glioblastoma survival were investigated. While MDM2 over-expression in glioblastoma is reported more frequently in the short-term survivors (≤1.5 years) than in the long-term (≥3 years) [8], we found that MDM2 was down-regulated in intermediate glioblastomas (12-36 months) compared to short-term glioblastomas (<12 months), and this is supported by previous research [8].

While the overall survival of EGFR-amplified glioblastomas is significantly shorter than controls (13.3 months vs. 26.6, p-value = 0.005) [4], EGFR was up-regulated in intermediate glioblastomas (vs. short-term glioblastomas) and down-regulated in long-term glioblastomas (vs. intermediate glioblastomas) and long-term glioblastomas compared to short-term glioblastomas. Potential reasons are the influ-
ence of other genetic alterations, such as TERT promoter mutation effect on survival time [4].

EZH2 over-expression in glioblastoma is an independent and negative predictor of overall survival [7]. Herein, EZH2 was down-regulated in long-term glioblastomas (>36 months) compared to intermediate glioblastomas (12–36 months) and this is consistent with previous research [7]; thus confirming the efficiency and accuracy of our analytical methods.

Although there is scant knowledge of associations between LINGO1/VEGFA/C7orf31 and glioblastoma survival, LINGO1 codes leucine-rich repeat and immunoglobulin domain-containing Nogo receptor-interacting protein 1. LINGO1 is a potent promoter for the maturation of neural stem cells to neurons [15], and this positively correlates with good prognosis for glioblastomas. LINGO1 is a potential biomarker highly expressed on the glioma surface; it is positive in >70% low-grade gliomas and its expression in
grade IV gliomas is significantly less than that in grades II and III [16]. Further, it may mediate the PHEN suppression in proliferation and invasion of glioma cells [17], and high grade glioma and proliferation/invasion generally lead to short survival and also signal poor prognosis for glioblastomas.

Herein, LINGO1 expression was up-regulated in intermediate glioblastomas (vs. short-term glioblastomas), and further up-regulated in long-term glioblastomas (vs. intermediate glioblastomas), and we established further that it is a long-term survival-related DEG and is down-regulated in glioblastoma compared to normal brain tissue. The agreement of our results with previous studies confirms that LINGO1 promotes good prognosis in glioblastoma [15–17].

Although little is known about C7orf31 function, it codes for chromosome 7 open reading frame 31 and we established that it is down-regulated in intermediate glioblastomas (vs. short-term glioblastomas).
short-term glioblastomas), and further down-regulated in long-term glioblastomas (vs. intermediate glioblastomas). Although it is accepted as a commonly down-regulated and short-term survival-related DEG, it was up-regulated in glioblastoma compared to normal brain cells. This suggests C7orf31 provides poor glioblastoma prognosis.

The VEGFA gene codes for vascular endothelial growth factor α, it enhances endothelial cells mitogenesis and permeability and is essential for cancer neo-vascularization. VEGFA can induce the development of glioblastoma in an implantation model of intracranial glioblastoma [19] and in vivo antibody-mediated suppression of VEGF-induced neo-vascularization inhibits tumor growth [18]. Injection of VEGF-expressing cancer stem cells in mouse brains causes massive glioblastoma expansion and high morbidity [20]. All these results indicate that VEGFA promotes develop-
ment and expansion of glioblastoma and subsequent short-term survival. We found that VEGFA was significantly down-regulated in long-term glioblastomas (vs. short-term and intermediate glioblastomas) and that it was a short-term survival-related DEG up-regulated in glioblastoma compared to normal brain tissue. This agrees with previous studies which indicate that VEGFA induces poor glioblastoma prognosis [19, 20].

Conclusion

In summary, gene expression analysis in primary glioblastomas with short-term, intermediate, and long-term survivals determined LINGO1, C7orf31 and VEGFA as prognostic genes of primary glioblastoma. Although their prognostic capacity requires further validation, our findings provide directions for the development of better targeted therapy for primary glioblastoma.

References


PROGNOSTIC GENES OF PRIMARY GLIOBLASTOMA 541