A thirteen-gene set efficiently predicts the prognosis of glioblastoma

HUYIN YANG, LUHAO JIN and XIAOYANG SUN

Department of Neurosurgery, Affiliated Huaian No. 1 People's Hospital of Nanjing Medical University,
Huai'an, Jiangsu 223300, P.R. China

Abstract. Glioblastoma multiforme (GBM) is the most common type of brain cancer; it usually recurs and patients have a short survival time. The present study aimed to construct a gene expression classifier and to screen key genes associated with GBM prognosis. GSE7696 microarray data set included samples from 10 recurrent GBM tissues, 70 primary GBM tissues and 4 normal brain tissues. Seed genes were identified by the 'survival' package in R and subjected to pathway enrichment analysis. Prognostic genes were selected from the seed genes using the 'rbsurv' package in R, unsupervised hierarchical clustering, survival analysis and enrichment analysis. Multivariate survival analysis was performed for the prognostic genes, and the GBM data set from The Cancer Genome Atlas database was utilized to validate the prognostic genes. Of the 1,785 seed genes analyzed, 13 prognostic feature genes, including collagen type XXVIII α1 chain (*COL28A1*), PDS5 cohesin-associated factor A (*PDS5A*), zinc-finger DHHC-type containing 2 (*ZDHHC2*), zinc-finger protein 24 (*ZNF24*), myosin VA (*MYO5A*) and myeloid/lymphoid or mixed-lineage leukemia translocated to 4 (*MLLT4*), were identified. These genes performed well on sample classification and prognostic risk differentiation, and six pathways, including adherens junction, cyclic adenosine 3',5'-monophosphate signaling and Ras signaling pathways, were enriched for these feature genes. The high-risk group was slightly older compared with the low-risk group. The validation data set confirmed the prognostic value of the 13 feature genes for GBM; of these, *COL28A1*, *PDS5A*, *ZDHHC2*, *ZNF24*, *MYO5A* and *MLLT4* may be crucial. These results may aid the understanding of the pathogenesis of GBM and provide important clues for the development of novel diagnostic markers or therapeutic targets.

Correspondence to: Dr Xiaoyang Sun, Department of Neurosurgery, Affiliated Huaian No. 1 People's Hospital of Nanjing Medical University, 6 Beijing West Road, Huai'an, Jiangsu 223300, P.R. China
E-mail: hayanghuyin@sina.com

Introduction

Glioblastoma multiforme (GBM) is the most common and the most invasive subtype of brain cancer; it is characterized by symptoms that include personality changes, headaches, nausea and unconsciousness (1,2). GBM originates from normal brain cells or low-grade astrocytoma, and may be induced by genetic disorders and radiation exposure (3,4). Clinical techniques for the treatment of GBM include surgery combined with radiation therapy or chemotherapy; however, the survival benefit is limited to ~12-15 months, or even shorter if the disease recurs (4).

Gene therapy is a novel strategy for treating cancers (5). Transient receptor potential genes are overexpressed in GBM, which promote the survival of patients (6,7). It has been previously reported that low expression of B cell-specific Moloney murine leukemia virus integration site 1 suppresses proliferation and promotes apoptosis of U251 GBM cells, and enhances the chemosensitivity of these cells to cisplatin (8,9). The expression level of epidermal growth factor-containing fibulin extracellular matrix protein 1 (*EFEMP1*) was associated with the survival of patients with GBM treated with temozolomide (TMZ) (10); thus, *EFEMP1* is considered a target for overcoming TMZ-resistance in GBM. Enhancer of zeste homolog 2 (*EZH2*) overexpression was associated with tumor grade and predicts short overall survival in patients with GBM (11); thus, *EZH2* may be a promising prognostic factor and therapeutic target for patients. Additionally, HOX transcript antisense RNA (*HOTAIR*) overexpression was associated with poor outcome in patients with GBM, and *HOTAIR* may be a therapeutic molecular target for this disease (12,13). O-6-Methylguanine-DNA methyltransferase (*MGMT*) methylation status and mutations in the isocitrate dehydrogenase 1 (*IDH1*) gene are two known clinicopathological factors linked to overall survival of patients with GBM (14). *MGMT* methylation is significantly associated with the clinical prognosis of GBM (15); *IDH1* is a prognostic marker of GBM, and mutations in this gene diminish the malignant progression of glioma (16). In addition, high expression of interleukin-13R mRNA is strongly associated with poor prognosis of GBM (17). POZ/BTB and AT hook-containing zinc-finger 1 is another prognostic marker of GBM; it is overexpressed in GBM-derived glioma-initiating stem cells, and is associated with the characteristic stem cell capacity to grow as

neurospheres *in vitro* (18). Despite this collective knowledge, the genes involved in the prognosis of GBM have not been comprehensively reported.

In 2008, Murat *et al* (19) established the GSE7696 gene expression profile, and demonstrated that high expression levels of epidermal growth factor receptor and stem cell-related 'self-renewal' signature are involved in the resistance to concomitant chemoradiotherapy of GBM. In 2011, Lambiv *et al* (20) used the GSE7696 data set to explore the action mechanism of tumor suppressor gene Wnt inhibitory factor 1 (*WIF1*) in GBM models, and concluded that WIF1 may have a tumor suppressive role in GBM through senescence.

However, this data set has not been fully explored. Using the GSE7696 data, additional key genes associated with the prognosis of GBM were investigated using comprehensive bioinformatics methods, such as survival analysis, enrichment analysis and hierarchical clustering. Results from the present study provided novel insights into the prognosis of GBM and may aid in the development of novel therapeutic approaches.

Materials and methods

Data source. The GSE7696 microarray data set, based on the GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array (Thermo Fisher Scientific, Inc.) platform, along with the corresponding clinical information was downloaded from the Gene Expression Omnibus database (www.ncbi.nlm.nih.gov/geo). The data set included 10 recurrent GBM tissue samples (2 females and 8 males; mean age, 51.31 years), 70 primary GBM tissue samples (19 females and 51 males; mean age, 48.07 years) and 4 normal brain tissue samples (sex and age information not available). All patients participated in a phase II or randomized phase III trial (21,22), with informed consent provided. The GSE7696 data set is available and the study was approved by the local ethics committee (19).

Selection of probes with expression changes in primary GBM tissue samples. The normalized data of GSE7696 were obtained, and the primary GBM tissue samples data were selected to use in the present study. Firstly, the unloaded probes were removed. Subsequently, the probes with varied expressions among 70 different patients with primary GBM were identified by the following steps. Firstly, the variance in probe expression level in each sample was calculated, and probes with variance <20% of the total probe variance were excluded. Secondly, the median probe expression in each sample was calculated and, as in the prior step, probes with a value <20% of the total median probe expression were removed. Finally, the probes with expression changes among these primary GBM tissue samples were selected.

Survival analysis and pathway enrichment analysis. Using the 'survival' package (23) in R (www.r-project.org), univariate survival analysis was conducted for the above probes with expression changes, and those with P<0.05 were considered as seed genes. The Kyoto Encyclopedia of Genes and Genomes (KEGG; www.genome.ad.jp/kegg) is a database used to search for gene functions, connecting genomic with functional information (24). Using the DAVID database (david.abcc.ncifcrf.gov) (25), KEGG pathway enrichment analysis was performed

for the seed genes to explore their functions. The categories with P<0.05 were considered to indicate a statistically significant difference.

Screening of prognostic feature genes, unsupervised hierarchical clustering and analysis of prognostic characteristics. Using the 'rbsurv' package in R (bioconductor.org/packages/release/bioc/html/rbsurv.html) (26), robust likelihood-based survival modeling was constructed to identify prognostic feature genes (27). The samples were classified based on the expression profiles of the prognostic feature genes using unsupervised hierarchical clustering (28). Thereafter, the prognostic differences among the classified samples were analyzed by the Kaplan-Meier survival analysis (29). The expression differences of the feature genes between primary GBM and normal samples were analyzed; and the scatter plot of the gene expression levels were drawn using the corrplot package (https://cran.r-project.org/web/packages/corrplot/vignettes/corrplot-intro.html) in R (version 3.4.4).

Functional and pathway enrichment analysis for the feature genes. The Gene Ontology (GO) database (www.geneontology.org) is used to predict potential functions of genes and their products (30). The prognostic feature genes were examined with GO functional and KEGG pathway enrichment analysis by the 'clusterProfiler' package in R (bioconductor.org/packages/release/bioc/html/clusterProfiler.html), with the threshold q-value <0.05 (31).

Multivariate survival analysis. The prognostic feature genes underwent multivariate survival analysis to check their overall influences on prognosis. In addition, the 'survivalROC' package in R (cran.r-project.org/web/packages/survivalROC/index.html) was used to draw the receiver operating characteristic (ROC) curve and calculate the area under the ROC curve (AUC) (32). Specifically, the survivalROC package was used to calculate the 'true positive rate' and 'false positive rate' of each sample, and the differences between the 'true positive rate' and 'false positive rate' for each sample were subsequently calculated. The sample with the smallest difference value was set as the cut-off in the multifactorial cox regression analysis, and samples with a higher value than this cut-off were deemed as the high-risk group, while those with a lower value were the low-risk group.

Validation of the prognostic feature genes using an independent data set. The GBM dataset in The Cancer Genome Atlas (TCGA; cancergenome.nih.gov) database (downloaded in January 27, 2015; based on the Illumina HiSeq platform, and included RNA-sequencing data in level 3 and clinical follow-up data) was obtained to validate the prognostic feature genes. The 172 samples in the data set comprised 13 recurrent GBM samples, 154 primary GBM samples and 5 adjacent normal tissue samples. Of the 154 primary GBM samples, 2 samples had no survival information. Finally, 152 primary GBM samples were selected for analysis. The data were transformed using $\log_2(x+1)$, followed by Cox regression analysis using the 'survival' package in R to compare the differences of prognosis and recurrence among the samples in different
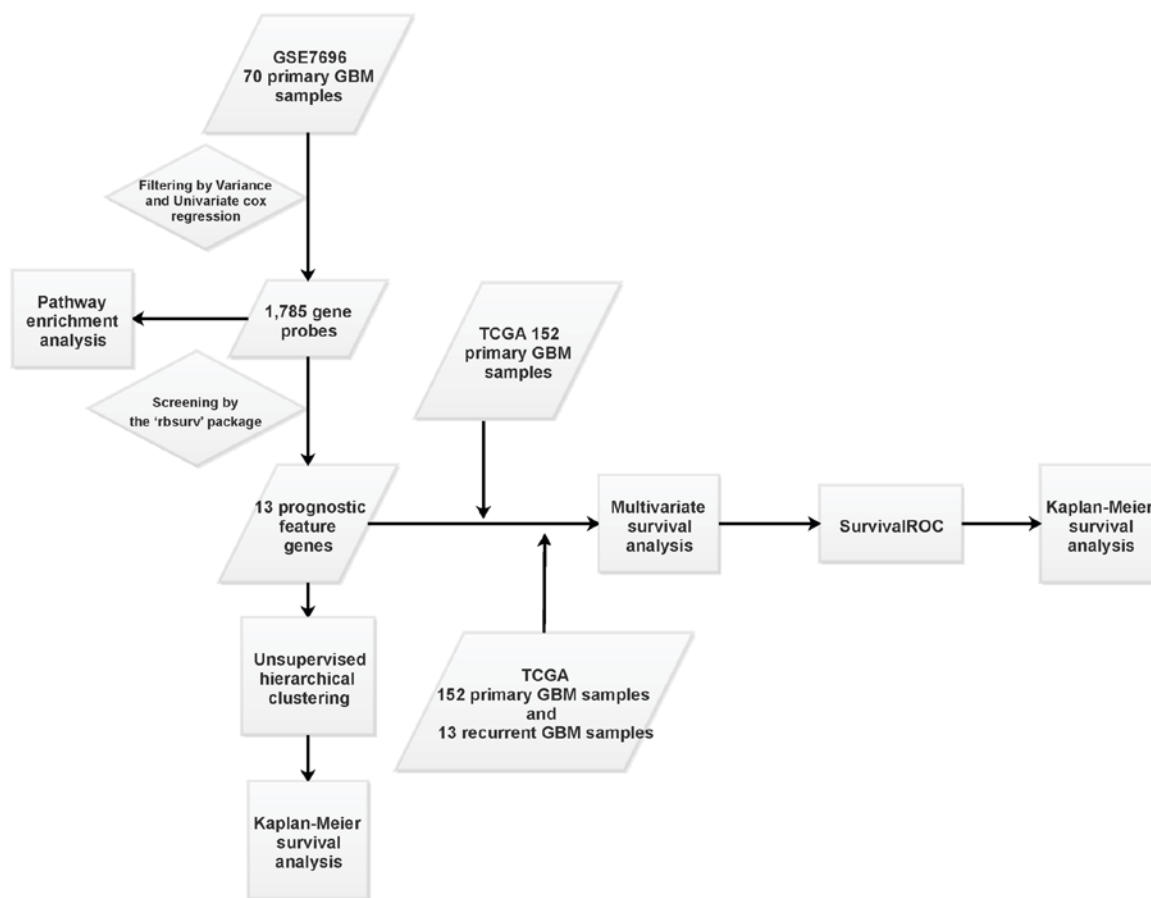
Figure 1. Bioinformatics workflow. The GSE7696 dataset that comprised 70 primary GBM samples was downloaded from the GEO database. Upon filtering out the unloaded probes, a total of 1,785 probes were obtained; the corresponding genes were subjected to pathway enrichment analysis. A total of 13 prognostic feature genes were screened and subjected to unsupervised hierarchical clustering analysis and KM survival analysis. The TCGA dataset that included 152 primary GBM samples was used to validate the results of the 13 prognostic feature genes, and was subjected to multivariate survival analysis and KM survival analysis. GBM, glioblastoma multiforme; GEO, Gene Expression Omnibus; KM, Kaplan-Meier; TCGA, The Cancer Genome Atlas.

groups. Differences in the identified factors (age and sex distribution) were assessed using R software. The Mann-Whitney test analysis was performed using the respective internal function in R, 'wilcox.test'. The χ^2 value was calculated using the 'chisq.test' function in R.

Results

Survival analysis and pathway enrichment analysis for seed genes. The bioinformatics analysis process used in the present study is outlined in Fig. 1. Following calculation of the variance and median for the probes in the 70 primary GBM, a total of 38,370 probes with expression changes were identified. Based on the univariate survival analysis using the 'survival' package, a total of 1,785 gene probes significantly correlated with prognosis of GBM were selected as seed genes. KEGG pathway enrichment analysis identified six pathways that were significantly enriched for these seed genes, including 'ribosome', 'regulation of actin cytoskeleton', 'endometrial cancer', 'pathways in cancer', 'amino sugar and nucleotide sugar metabolism' and 'non-small cell lung cancer' (Table I).

Screening of prognostic feature genes, unsupervised hierarchical clustering and analysis of prognostic characteristics. Robust likelihood-based survival modeling

identified 13 prognostic feature genes from the 1,785 seed genes, including collagen type XXVIII α1 chain (*COL28A1*), PDS5 cohesin-associated factor A (*PDS5A*; also known as sister chromatid cohesion protein 112), zinc-finger DHHC-type containing 2 (*ZDHHC2*), zinc-finger protein 24 (*ZNF24*), myosin VA (*MYO5A*) and myeloid/lymphoid or mixed-lineage leukemia translocated to 4 (*MLLT4*) (Table II). Unsupervised hierarchical clustering was conducted for these 13 prognostic feature genes. A heat map demonstrated that these genes may be used to classify the 70 primary GBM samples into two clusters: Cluster 1 contained 51 samples and Cluster 2 contained 19 samples (Fig. 2). Subsequently, Kaplan-Meier survival analysis was used to analyze the prognostic differences between the two clusters. Patients in the Cluster 1 and in Cluster 2 had significant differences in their prognosis, which indicated that the 13 prognostic feature genes may effectively differentiate between high-risk and low-risk patients in a clinical setting (Fig. 3A). In addition, the expression levels of the 13 prognostic feature genes were relatively low, which suggested that they had low redundancy (Fig. 3B).

Functional annotation of prognostic feature genes. Using the 'clusterProfiler' package, functional annotation of the prognostic feature genes were conducted. The results identified three genes, including, natriuretic peptide receptor

Table I. Pathways significantly enriched for the 1,785 seed genes.

Term	Count	P-value
hsa03010:Ribosome	43	2.31×10^{-26}
hsa04810:Regulation of actin cytoskeleton	27	4.09×10^{-03}
hsa05213:Endometrial cancer	10	9.29×10^{-03}
hsa05200:Pathways in cancer	34	2.08×10^{-02}
hsa00520:Amino sugar and nucleotide sugar metabolism	8	3.19×10^{-02}
hsa05223:Non-small cell lung cancer	9	3.31×10^{-02}

3 (*NPR3*), ANKH inorganic pyrophosphate transport regulator and OTU deubiquitinase, ubiquitin aldehyde binding 1 (*OTUB1*) were significantly enriched in functions, including 'natriuretic peptide receptor activity', 'inorganic phosphate transmembrane transporter activity' and 'NEDD8-specific protease activity', respectively. In addition, myeloid/lymphoid or mixed-lineage leukemia translocated to 4 (*MLLT4*; also known as afadin, adherens junction formation factor) was indicated to be involved in six KEGG pathways, including 'adherens junction', 'leukocyte transendothelial migration', 'tight junction', 'cAMP signaling pathway', 'Rap1 signaling pathway' and 'Ras signaling pathway'. *COL28A1*, *PDS5A*, *ZDHHC2* and *MYO5A* were enriched in 'collagen biosynthesis and modifying enzymes', 'separation of sister chromatids', 'surfactant metabolism' and 'regulation of actin dynamics for phagocytic cup formation' pathways, respectively.

Multivariate survival analysis. To check the overall influence of the prognostic feature genes on prognosis, multivariate survival analysis was carried out. These genes were classified as having a positive effect on prognosis (area under the ROC curve=0.826) for the 70 primary GBM samples (Fig. 4A). Furthermore, Kaplan-Meier survival analysis demonstrated that the samples in the high-risk and low-risk groups differed significantly in their prognosis ($P=5.67 \times 10^{-12}$; Fig. 4B), which indicated that the 13 prognostic feature genes may effectively distinguish samples with a different prognostic risk.

Validation of the prognostic feature genes using another independent data set. The GBM data set was downloaded from the TCGA database and used to validate the prognostic feature genes. Survival analysis demonstrated that the 13 prognostic feature genes had good classification effects on samples in the validation data set (area under the ROC curve =0.886) (Fig. 5A). The high-risk and low-risk groups differed significantly in their prognosis ($P=5.00 \times 10^{-5}$; Fig. 5B). Furthermore, age distribution analysis of the classified primary GBM samples demonstrated that samples in the high-risk and low-risk groups differed significantly in age, with the high-risk group being older than the low-risk group (61.54 vs. 56.96; P=0.04396; Table III). There was no significant difference concerning sex distribution of the classified samples.

Table II. A total of 13 prognostic feature genes were identified from the 1,785 seed genes.

Gene ID	nloglik	AIC	Gene
219789_at	198.94	399.87	*NPR3*
215201_at	193.8	391.6	*REPS1*
213254_at	190.47	386.95	*TNRC6B*
225526_at	186.37	380.73	*MKLN1*
223093_at	183.88	377.77	*ANKH*
239921_at	180.76	373.52	*COL28A1*
217331_at	177.81	369.62	*PDS5A*
224685_at	177.78	371.55	*MLLT4*
244779_at	176.76	371.52	*ZDHHC2*
228786_at	176.17	372.34	*PTCHD3P1*
201245_s_at	175.81	373.63	*OTUB1*
1554045_at	173.56	371.12	*ZNF24*
204527_at	170.83	367.67	*MYO5A*

AIC, Akaike's information criterion; *ANKH*, ANKH inorganic pyrophosphate transport regulator; *COL28A1*, collagen type XXVIII α1 chain; ID, identification; *MKLN1*, muskelin 1; *MLLT4*, myeloid/lymphoid or mixed-lineage leukemia translocated to 4; *MYO5A*, myosin VA; *NPR3*, natriuretic peptide receptor 3; *OTUB1*, OTU deubiquitinase, ubiquitin aldehyde binding 1; *PDS5A*, PDS5 cohesin-associated factor A; *PTCHD3P1*, patched domain-containing 3 pseudogene 1; *REPS1*, PALBP1-associated EPS domain-containing 1; *TNRC6B*, trinucleotide repeat-containing 6B; *ZDHHC2*, zinc-finger DHHC-type containing 2; *ZNF24*, zinc-finger protein 24.

Discussion

In the present study, a total of 1,785 gene probes with significant prognostic differences were selected as seed genes. From these, 13 prognostic feature genes, including *COL28A1*, *PDS5A*, *ZDHHC2*, *ZNF24*, *MYO5A*, *MLLT4*, *NPR3*, *ANKH* and *OTUB1*, were further screened. The prognostic feature genes performed well for the classification of samples into different prognostic risk categories. Additionally, the GBM data set downloaded from the TCGA database further confirmed the association of the 13 prognostic feature genes with the prognosis of GBM.

Collagen XVI may promote the invasion of glioma cells by damaging cell-cell interactions or mediating the β1-integrin activation pattern, which provides new approaches for treating cancer in neuro-oncology (33). Fibrillar collagens and the collagen internalization receptor, endocytic receptor 180 (Endo180), are overexpressed in GBM (34). Additionally, Endo180 affects the invasion and progression of tumors (34). Collagen type I α1 (*CO1A1*) was reported to serve a suppressive biological role in the progression of glioma, suggesting that it may be applied to treating the disease (35). The COL28A1 protein is a filament-forming collagen and is detected in the adult sciatic nerve (36). In our study, *COL28A1* was enriched in 'collagen biosynthesis and modifying enzymes' pathway. Therefore, *COL28A1* may serve a role in the progression of GBM.

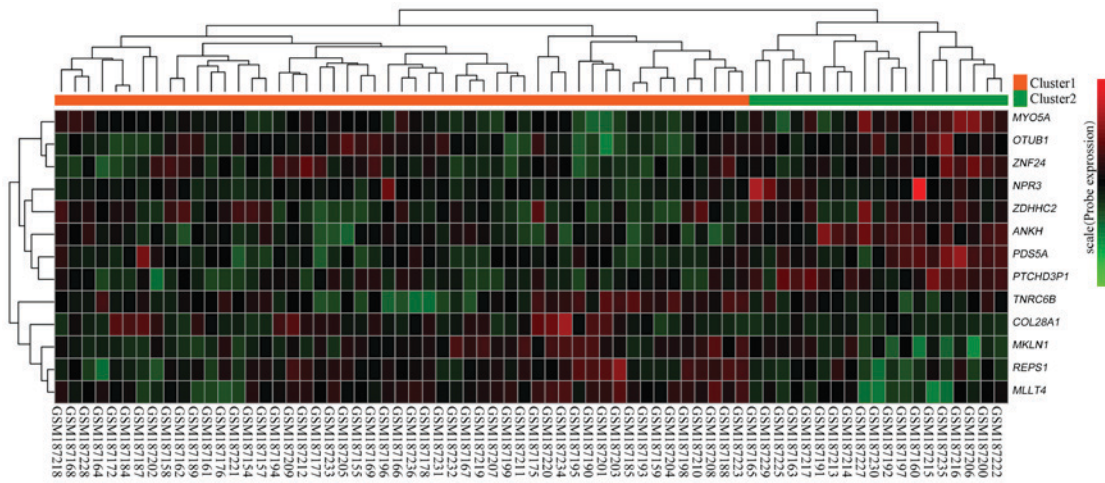*PDS5A* is tissue-specific and has two-fold effects in tumorigenesis, acting as a tumor suppressor or an oncogenic

Figure 2. Clustering heatmap for the 13 prognostic feature genes. The horizontal axis represents the samples, and the vertical axis represents the genes.
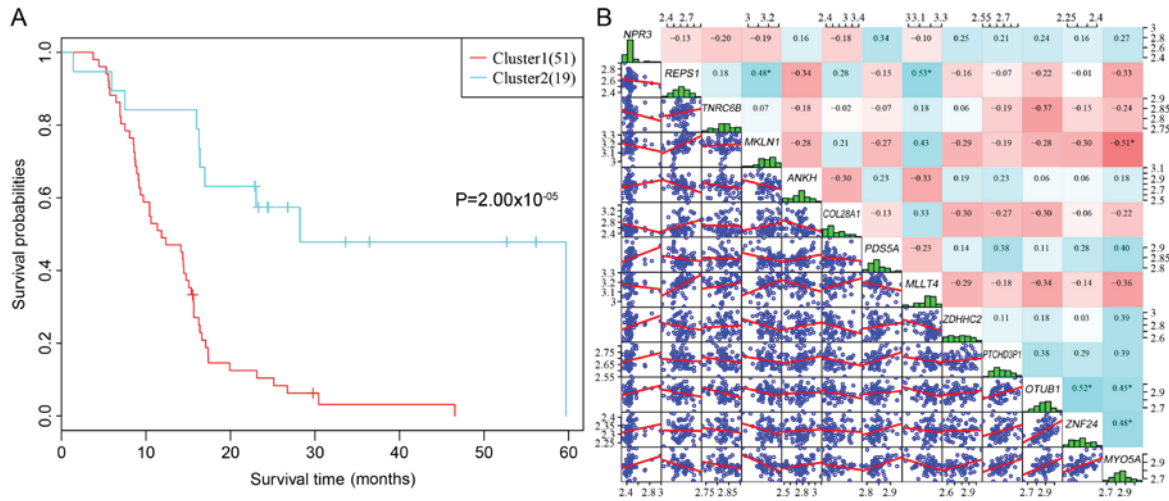

Figure 3. Kaplan-Meier survival curve and expression correlations analysis for the 13 prognostic feature genes. (A) The survival curve of the two clusters. (B) The lower-left part is the scatter plot of the gene expression levels. The red to blue color in the upper-right part represent correlation coefficients ranging from -1 to +1, respectively. The diagonal line represents the expression distribution histogram of each gene. Asterisks indicate a correlation coefficient ≥0.45.
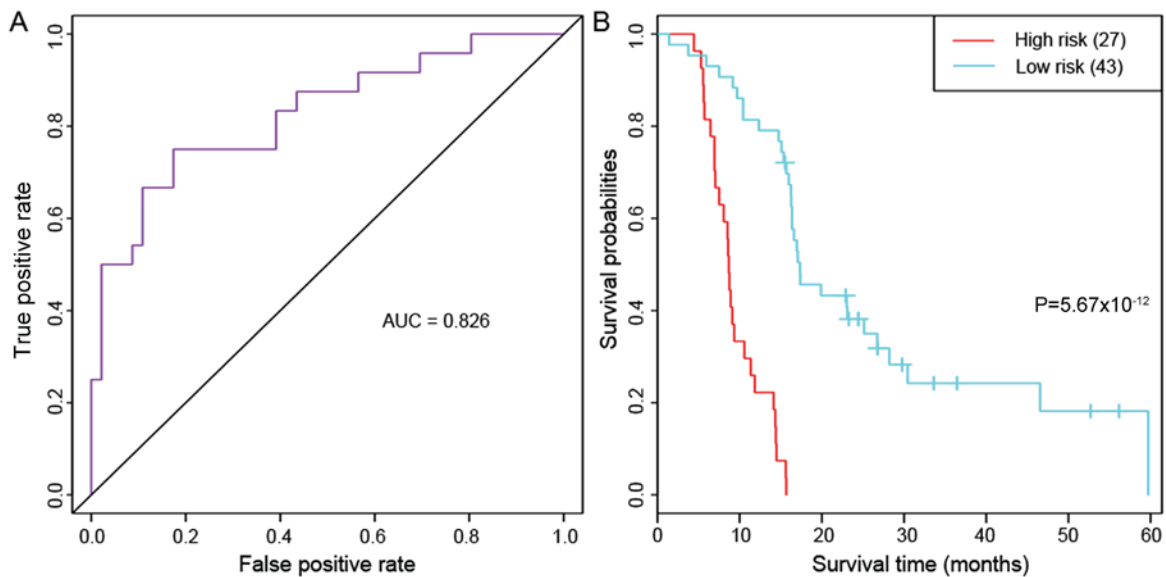

Figure 4. Multivariate survival analysis for the 13 prognostic feature genes in the 70 primary GBM samples in GSE7696. (A) The AUC of multivariate survival analysis. (B) The Kaplan-Meier survival curve. AUC, area under the Receiver Operating Characteristic curve.

Table III. Age distribution of the samples in the high-risk and low-risk groups.

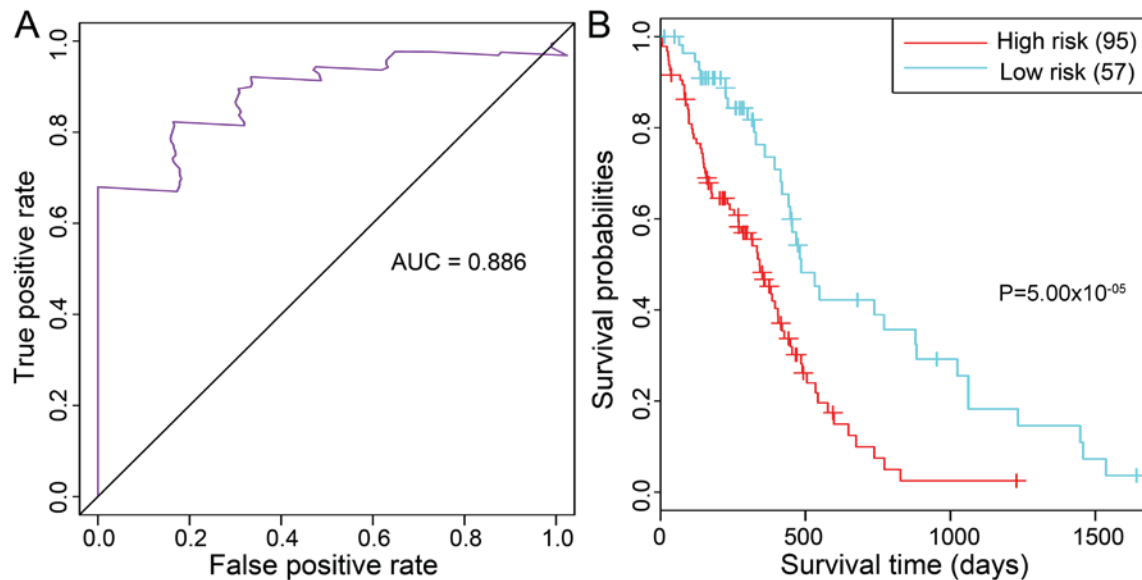| Sample group	Min age	Mean age	Max age	Mann-Whitney test	Female	Male	P-value
High risk (n=95)	21	61.54	89	P=0.04396	32	64	P=0.51
Low risk (n=57)	21	56.96	85		22	35	

Figure 5. Multivariate survival analysis for the prognostic feature genes in the 152 primary GBM samples in TCGA (validation data set). (A) The AUC multivariate survival analysis. (B) The Kaplan-Meier survival curve. AUC, area under the Receiver Operating Characteristic curve.

factor to promote tumor proliferation (37). *PDS5A* was previously reported to be overexpressed in high-grade gliomas and is positively associated with the World Health Organization grade of gliomas (37). Based on our enrichment result, *PDS5A* was involved in 'separation of sister chromatids' pathway, indicating it might influence the tumor's DNA synthesis phase in GBM development.

ZDHHC2 on chromosome 8p21.3-22 belongs to the DHHC-domain protein family of protein acyltransferases (38). As the largest transcription factor family in humans, zinc-finger proteins serve roles in multiple biological processes, including autophagy, metabolism, development and differentiation, and in cancer progression (39). In addition, *ZDHHC2* is associated with prognosis in a number of types of cancer. In gastric cancer, decreased *ZDHHC2* expression is associated with lymph node metastasis and indicates a poor prognosis (40). In the present study, *ZDHHC2* was highly associated with the 'surfactant metabolism' pathway and identified as a prognostic gene in GBM, suggesting that it might have an important role in GBM prognosis.

ZNF24 plays a critical role in brain development and various cancer types. Furthermore, *ZNF24* may contribute to cell cycle promotion and maintenance of the progenitor stage of neural cells (41). Platelet-derived growth factor receptor (PDGFR) signaling is a crucial mechanism for the initiation and development of GBM (42). In GBM cell lines, *ZNF24* negatively regulates two transcription factors, vascular endothelial growth factor (VEGF) and PDGFR-β (43).

In the present study, the above three genes were among the 13 crucial prognostic genes of GBM. The collective findings implicate *PDS5A*, *ZDHHC2* and *ZNF24* are involved in the pathogenesis of GBM. Although no studies have reported the potential roles of *ZDHHC2* in GBM, this gene may also be associated with the prognosis of GBM, based on our results.

Myosin II is required for the invasion of glioma cells and is a promising target for the anti-invasive treatment of malignant brain tumors (44,45). In 1321N1 GBM cells, *MYO1C* is crucial for lamellipodia formation to produce a protein complex promoting cell migration (46). *MYO6* was implicated in human glioma and its inhibition may be a promising therapeutic method for the disease (47). The activation of myosin-associated contractility sensitizes GBM tumor-initiating cells, which subsequently weakens the invasive ability of the cells (48). In the present study, *MYO5A* was enriched in 'regulation of actin dynamics for phagocytic cup formation'. Thus, *MYO5A* may serve a role in the progression of GBM.

MLLT4 is one of the aliases of the gene, afadin, adherens junction formation factor, which belongs to an adhesion system and encodes a protein participating in cell junctions during embryogenesis. Neurofibromin 1 (*NF1*) mutations have been identified in several cancer types, including GBM (49). Furthermore, in a mouse model of *NF1*-associated optic pathway gliomas (OPG), cyclic adenosine 3',5'-monophosphate (cAMP) inhibits the growth of OPG (50). Notably, *MLLT4* is one of the *NF1*-regulated effectors downstream of RAS (51). Results from the present studies support the possible

importance of *MLLT4* in GBM pathogenesis through the participation in the adherens junction, cAMP signaling and Ras signaling pathways. In addition, a number of pathways, such as leukocyte transendothelial migration, are enriched in GBM-derived extracellular vesicles (EVs), which suggested that GBM cells may exhibit mechanisms to selectively combine these proteins in EVs (52). In the present study, leukocyte transendothelial migration was one of the six enriched pathways identified for MLLT3; therefore, this pathway may also be an important regulation target in GBM development. With regard to the tight junction pathway, junctional adhesion molecule (JAM) is one family of the immunoglobulin-like superfamily expressed in tight junctions, and abnormal JAM-A expression was reported to contribute to the progression of GBM (53). This may be evidence of the gene expression alterations in the tight junction pathway in the development of GBM.

NPR3 encodes a natriuretic peptide receptor, which regulates metabolic processes. A long noncoding RNA (lncRNA), BCYRN1, has an oncogenic role in colorectal cancer cells by upregulating *NPR3* expression (54). In clear cell renal cell carcinoma (ccRCC), the lncRNA MRCCAT1 has been observed to promote ccRCC metastasis by inhibiting *NPR3* (55). *ANKH* is identified as a novel putative oncogene in small cell lung cancer cell lines (56). In cervical cancer, *ANKH* is also suggested as an oncogene and the upregulation is validated by reverse transcription-quantitative polymerase chain reaction (57). To the best of our knowledge, there have been no studies reporting an association between these two genes, *NPR3* and *ANKH*, and glioma. However, in the present study, these were important prognostic genes of GBM and were enriched in crucial pathways, suggesting that they may be novel markers in GBM and correlate with prognosis.

The OTUB1 protein is a deubiquitinating enzyme. In glioma tissues, expression of *OTUB1* is increased and the expression level is associated with the glioma grade; on the other hand, knockdown of *OTUB1* suppresses the tumor cell migration (58). This finding suggests that *OTUB1* serves an important role in the etiology of glioma (58).

However, all these predictive results need to be further validated by *in vitro* and *in vivo* experiments. Notably, patients in the high-risk group were slightly older compared with patients in the low-risk group; although the difference was minor, it was significant. A previous study identified age as a risk factor for the process of dexamethasone-induced leukocytosis, which is associated with poor survival in the newly diagnosed GBM (59). Therefore, the present study suggested that age may be a risk factor of GBM, but this needs to be validated using larger data sets.

In conclusion, the 13-gene set was tested and verified, and very efficiently predicted the prognosis of GBM in independent data sets. In addition, *COL28A1*, *PDS5A*, *ZDHHC2*, *ZNF24*, *MYO5A* and *MLLT4* were implicated as key genes involved in the prognosis of GBM. The adherens junction, cAMP and Ras signaling pathways may be important in the progression of GBM, and age may be a risk factor for prognosis.

Acknowledgements

Funding

Availability of data and materials

The analyzed data sets generated during the present study are available from the corresponding author on reasonable request.

Authors' contributions

HY performed data analyses and wrote the manuscript. LJ contributed significantly to the data analysis and manuscript revision. XS conceived and designed the study. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Patient consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Bleeker FE, Molenaar RJ and Leenstra S: Recent advances in the molecular understanding of glioblastoma. J Neurooncol 108: 11-27, 2012.
2. Young RM, Jamshidi A, Davis G and Sherman JH: Current trends in the surgical management and treatment of adult glioblastoma. Ann Transl Med 3: 121, 2015.
3. McGuire S: World Cancer Report 2014. Geneva, Switzerland: World Health Organization, International Agency for Research on Cancer, WHO Press, 2015. Adv Nutr 7: 418-419, 2016.
4. Gallego O: Nonsurgical treatment of recurrent glioblastoma. Curr Oncol 22: e273-e281, 2015.
5. Fulci G and Chiocca EA: The status of gene therapy for brain tumors. Expert Opin Biol Ther 7: 197-208, 2007.
6. Alptekin M, Eroglu S, Tutar E, Sencan S, Geyik MA, Ulasli M, Demiryurek AT and Camci C: Gene expressions of TRP channels in glioblastoma multiforme and relation with survival. Tumor Biology 36: 9209-9213, 2015.
7. Chen J, Luan Y, Yu R, Zhang Z, Zhang J and Wang W: Transient receptor potential (TRP) channels, promising potential diagnostic and therapeutic tools for cancer. Biosci Trends 8: 1-10, 2014.
8. Hong Y, Shang C, Xue YX and Liu YH: Silencing of Bmi-1 gene enhances chemotherapy sensitivity in human glioblastoma cells. Med Sci Monit 21: 1002-1007, 2015.
9. Ye L, Wang C, Yu G, Jiang Y, Sun D, Zhang Z, Yu X, Li X, Wei W, Liu P, *et al*: Bmi-1 induces radioresistance by suppressing senescence in human U87 glioma cells. Oncol Lett 8: 2601-2606, 2014.
10. Hiddingh L, Tannous BA, Teng J, Tops B, Jeuken J, Hulleman E, Boots-Sprenger SH, Vandertop WP, Noske DP, Kaspers GJ, *et al*: EFEMP1 induces γ-secretase/Notch-mediated temozolomide resistance in glioblastoma. Oncotarget 5: 363-374, 2014.
11. Zhang J, Chen L, Han L, Shi Z, Zhang J, Pu P and Kang C: EZH2 is a negative prognostic factor and exhibits pro-oncogenic activity in glioblastoma. Cancer Lett 356: 929-936, 2015.
12. Zhou X, Ren Y, Zhang J, Zhang C, Zhang K, Han L, Kong L, Wei J, Chen L, Yang J, *et al*: HOTAIR is a therapeutic target in glioblastoma. Oncotarget 6: 8353-8365, 2015.
13. Zhang K, Sun X, Zhou X, Han L, Chen L, Shi Z, Zhang A, Ye M, Wang Q, Liu C, *et al*: Long non-coding RNA HOTAIR promotes glioblastoma cell cycle progression in an EZH2 dependent manner. Oncotarget 6: 537-546, 2015.

14. Insin P and Prueksaritanond N: Evaluation of four risk of malignancy indices (RMI) in the preoperative diagnosis of ovarian malignancy at Rajavithi hospital. Thai J Obstet Gynaecol 21, 2013.

15. Sim J, Nam DH, Kim Y, Lee IH, Choi JW, Sa JK and Suh YL: Comparison of 1p and 19q status of glioblastoma by whole exome sequencing, array-comparative genomic hybridization, and fluorescence in situ hybridization. Med Oncol 35: 60, 2018.

16. Yao Q, Cai G, Yu Q, Shen J, Gu Z, Chen J, Shi W and Shi J: IDH1 mutation diminishes aggressive phenotype in glioma stem cells. Int J Oncol 52: 270-278, 2017.

17. Han J and Puri RK: Analysis of the cancer genome atlas (TCGA) database identifies an inverse relationship between interleukin-13 receptor α1 and α2 gene expression and poor prognosis and drug resistance in subjects with glioblastoma multiforme. J Neurooncol 136: 463-474, 2018.

18. Guadagno E, Vitiello M, Francesca P, Calì G, Caponnetto F, Cesselli D, Camorani S, Borrelli G, Califano M, Cappabianca P, *et al*: PATZ1 is a new prognostic marker of glioblastoma associated with the stem-like phenotype and enriched in the proneural subtype. Oncotarget 8: 59282-59300, 2017.

19. Murat A, Migliavacca E, Gorlia T, Lambiv WL, Shay T, Hamou MF, de Tribolet N, Regli L, Wick W, Kouwenhoven MC, *et al*: Stem cell-related 'self-renewal' signature and high epidermal growth factor receptor expression associated with resistance to concomitant chemoradiotherapy in glioblastoma. J Clin Oncol 26: 3015-3024, 2008.

20. Lambiv WL, Vassallo I, Delorenzi M, Shay T, Diserens AC, Misra A, Feuerstein B, Murat A, Migliavacca E, Hamou MF, *et al*: The Wnt inhibitory factor 1 (WIF1) is targeted in glioblastoma and has a tumor suppressing function potentially by induction of senescence. Neuro Oncol 13: 736-747, 2011.

21. Stupp R, Mason WP, van den Bent MJ, Weller M, Fisher B, Taphoorn MJ, Belanger K, Brandes AA, Marosi C, Bogdahn U, *et al*: Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. N Engl J Med 352: 987-996, 2005.

22. Stupp R, Dietrich PY, Ostermann Kraljevic S, Pica A, Maillard I, Maeder P, Meuli R, Janzer R, Pizzolato G, Miralbell R, *et al*: Promising survival for patients with newly diagnosed glioblastoma multiforme treated with concomitant radiation plus temozolomide followed by adjuvant temozolomide. J Clin Oncol 20: 1375-1382, 2002.

23. Therneau TM: Survival analysis (R package survival version 2.39-5). 2015.

24. Kanehisa M, Sato Y, Kawashima M, Furumichi M and Tanabe M: KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res 44: D457-D62, 2015.

25. Huang DW, Sherman BT, Tan Q, Collins JR, Alvord WG, Roayaei J, Stephens R, Baseler MW, Lane HC and Lempicki RA: The DAVID gene functional classification tool: A novel biological module-centric algorithm to functionally analyze large gene lists. Genome Boil 8: R183, 2007.

26. Cho HJ, Kim S, Kang J and Lee JW: How to use the rbsurv Package. 2010.

27. Renaud G, Stenzel U, Maricic T, Wiebe V and Kelso J: deML: Robust demultiplexing of Illumina sequences using a likelihood-based approach. Bioinformatics 31: 770-772, 2015.

28. Wang SJ: Unsupervised hierarchical clustering based on sequential partitioning and merging. International Symposium on Vlsi Design, Automation and Test, 2016.

29. Porcher R: CORR Insights(®): Kaplan-meier survival analysis overestimates the risk of revision arthroplasty: A meta-analysis. Clin Orthop Relat Res 473: 3431-3442, 2015.

30. Smith B, Williams J and Schulze-Kremer S: The ontology of the gene ontology. AMIA Annu Symp Proc: 609-613, 2003.

31. Yu G, Wang LG, Han Y and He QY: clusterProfiler: An R package for comparing biological themes among gene clusters. OMICS 16: 284-287, 2012.

32. Heagerty PJ, Thomas L and Pepe MS: Time-dependent ROC curves for censored survival data and a diagnostic marker. Biometrics 56: 337-344, 2000.

33. Bauer R, Ratzinger S, Wales L, Bosserhoff A, Senner V, Grifka J and Grässel S: Inhibition of collagen XVI expression reduces glioma cell invasiveness. Cell Physiol Biochem 27: 217-226, 2011.

34. Huijbers IJ, Iravani M, Popov S, Robertson D, Alsarraj S, Jones C and Isacke CM: A role for fibrillar collagen deposition and the collagen internalization receptor endo180 in glioma invasion. PLoS One 5: e9808, 2010.

35. Honma K, Miyata T and Ochiya T: Type I collagen gene suppresses tumor growth and invasion of malignant human glioma cells. Cancer Cell Int 7: 12, 2007.

36. Veit G, Kobbe B, Keene DR, Paulsson M, Koch M and Wagener R: Collagen XXVIII, a novel von Willebrand factor A domain-containing protein with many imperfections in the collagenous domain. J Biol Chem 281: 3494-3504, 2006.

37. Hagemann C, Weigelin B, Schommer S, Schulze M, Al-Jomah N, Anacker J, Gerngras S, Kühnel S, Kessler AF, Polat B, *et al*: The cohesin-interacting protein, precocious dissociation of sisters 5A/sister chromatid cohesion protein 112, is up-regulated in human astrocytic tumors. Int J Mol Med 27: 39-51, 2011.

38. Oyama T, Miyoshi Y, Koyama K, Nakagawa H, Yamori T, Ito T, Matsuda H, Arakawa H and Nakamura Y: Isolation of a novel gene on 8p21.3-22 whose expression is reduced significantly in human colorectal cancers with liver metastasis. Genes Chromosomes Cancer 29: 9-15, 2000.

39. Jen J and Wang YC: Zinc finger proteins in cancer progression. J Biomed Sci 23: 53, 2016.

40. Yan SM, Tang JJ, Huang CY, Xi SY, Huang MY, Liang JZ, Jiang YX, Li YH, Zhou ZW, Ernberg I, *et al*: Reduced expression of ZDHHC2 is associated with lymph node metastasis and poor prognosis in gastric adenocarcinoma. PLoS One 8: e56366, 2013.

41. Khalfallah O, Ravassard P, Che SL, Fligny C, Serre A, Bayard E, Faucon-Biguet N, Mallet J, Meloni R and Nardelli J: Zinc finger protein 191 (ZNF191/Zfp191) is necessary to maintain neural cells as cycling progenitor. Stem Cells 27: 1643-1653, 2009.

42. Carrasco-Garcia E, Martinez-Lacaci I, Mayor-López L, Tristante E, Carballo-Santana M, García-Morales P, Ventero Martin MP, Fuentes-Baile M, Rodriguez-Lescure Á and Saceda M: PDGFR and IGF-1R inhibitors induce a G2/M arrest and subsequent cell death in human glioblastoma cell lines. Cells 7: E131, 2018.

43. Li JZ, Chen X, Liu Y, Ding L, Qiu L, Hu ZL and Zhang J: The transcriptional repression of platelet-derived growth factor receptor-beta by the zinc finger transcription factor ZNF24. Biochem Biophys Res Commun 397: 318-322, 2010.

44. Beadle C, Assanah MC, Monzo P, Vallee R, Rosenfeld SS and Canoll P: The role of myosin II in glioma invasion of the brain. Mol Biol Cell 19: 3357-3368, 2008.

45. Ivkovic S, Beadle C, Noticewala S, Massey SC, Swanson KR, Toro LN, Bresnick AR, Canoll P and Rosenfeld SS: Direct inhibition of myosin II effectively blocks glioma invasion in the presence of multiple motogens. Mol Biol Cell 23: 533-542, 2011.

46. Edimo WE, Ramos AR, Ghosh S, Vanderwinden JM and Erneux C: The SHIP2 interactor Myo1c is required for cell migration in 1321 N1 glioblastoma cells. Biochem Biophys Res Commun 476: 508-514, 2016.

47. Xu R, Fang XH and Zhong P: Myosin VI contributes to malignant proliferation of human glioma cells. Korean J Physiol Pharmacol 20: 139-145, 2016.

48. Wong SY, Ulrich TA, Deleyrolle LP, MacKay JL, Lin JM, Martuscello RT, Jundi MA, Reynolds BA, Kumar S: Constitutive activation of myosin-dependent contractility sensitizes glioma tumor-initiating cells to mechanical inputs and reduces tissue invasion. Cancer Res 75: 1113-1122, 2015.

49. Brennan CW, Verhaak RG, McKenna A, Campos B, Noushmehr H, Salama SR, Zheng S, Chakravarty D, Sanborn JZ, Berman SH, *et al*: The somatic genomic landscape of glioblastoma. Cell 155: 462-477, 2013.

50. Warrington NM, Sun T, Luo J, Mckinstry RC, Parkin PC, Ganzhorn S, Spoljaric D, Albers AC, Merkelson A, Stewart DR, *et al*: The cyclic AMP pathway is a sex-specific modifier of glioma risk in type I neurofibromatosis patients. Cancer Res 75: 16-21, 2015.

51. Kiuru M and Busam KJ: The NF1 gene in tumor syndromes and melanoma. Lab Invest 97: 146-157, 2017.

52. de Vrij J, Maas SL, Kwappenberg KM, Schnoor R, Kleijn A, Dekker L, Luider TM, de Witte LD, Litjens M, van Strien ME, *et al*: Glioblastoma-derived extracellular vesicles modify the phenotype of monocytic cells. Int J Cancer 137: 1630-1642, 2015.

53. Leech AO, Cruz RG, Hill AD and Hopkins AM: Paradigms lost-an emerging role for over-expression of tight junction adhesion proteins in cancer pathogenesis. Ann Transl Med 3: 184, 2015.

54. Gu L, Lu LS, Zhou DL and Liu ZC: Long noncoding RNA BCYRN1 promotes the proliferation of colorectal cancer cells via Up-regulating NPR3 expression. Cell Physiol Biochem 48: 2337-2349, 2018.

55. Li JK, Chen C, Liu JY, Shi JZ, Liu SP, Liu B, Wu DS, Fang ZY, Bao Y, Jiang MM, *et al*: Long noncoding RNA MRCCAT1 promotes metastasis of clear cell renal cell carcinoma via inhibiting NPR3 and activating p38-MAPK signaling. Mol Cancer 16: 111, 2017.
56. Coe BP, Henderson LJ, Garnis C, Tsao MS, Gazdar AF, Minna, J, Lam S, Macaulay C and Lam WL: High-resolution chromosome arm 5p array CGH analysis of small cell lung carcinoma cell lines. Genes Chromosomes Cancer 42: 308-313, 2005.
57. Kloth JN, Oosting J, van Wezel T, Szuhai K, Knijnenburg J, Gorter A, Kenter GG, Fleuren GJ and Jordanova ES: Combined array-comparative genomic hybridization and single-nucleotide polymorphism-loss of heterozygosity analysis reveals complex genetic alterations in cervical cancer. BMC Genomics 8: 53, 2007.
58. Xu L, Li J, Bao Z, Xu P, Chang H, Wu J, Bei Y, Xia L, Wu P, Yan K, *et al*: Silencing of OTUB1 inhibits migration of human glioma cells in vitro. Neuropathology 37: 217-226, 2017.
59. Dubinski D, Won SY, Gessler F, Quick-Weller J, Behmanesh B, Bernatz S, Forster MT, Franz K, Plate KH, Seifert V, *et al*: Dexamethasone-induced leukocytosis is associated with poor survival in newly diagnosed glioblastoma. J Neurooncol 137: 503-510, 2018.