Clinical Cancer Research

A 4-Gene Signature Associated with Clinical Outcome in High-Grade Gliomas

Marie de Tayrac[1,2], Marc Aubry[1,2,3], Stephan Saïkali[4], Amandine Etcheverry[1,2,3], Cyrille Surbled[1], Frédérique Guénot[3], Marie-Dominique Galibert[1,2], Abderrahmane Hamlat[5], Thierry Lesimple[6], Véronique Quillien[7], Philippe Menei[8], and Jean Mosser[1,2,3]

Abstract

Purpose: Gene expression studies provide molecular insights improving the classification of patients with high-grade gliomas. We have developed a risk estimation strategy based on a combined analysis of gene expression data to search for robust biomarkers associated with outcome in these tumors.

Experimental Design: We performed a meta-analysis using 3 publicly available malignant gliomas microarray data sets (267 patients) to define the genes related to both glioma malignancy and patient outcome. These biomarkers were used to construct a risk-score equation based on a Cox proportional hazards model on a subset of 144 patients. External validations were performed on microarray data (59 patients) and on RT-qPCR data (194 patients). The risk-score model performances (discrimination and calibration) were evaluated and compared with that of clinical risk factors, *MGMT* promoter methylation status, and *IDH1* mutational status.

Results: This interstudy cross-validation approach allowed the identification of a 4-gene signature highly correlated to survival (*CHAF1B*, *PDLIM4*, *EDNRB*, and *HJURP*), from which an optimal survival model was built ($P < 0.001$ in training and validation sets). Multivariate analysis showed that the 4-gene risk score was strongly and independently associated with survival (hazard ratio = 0.46; 95% CI, 0.26–0.81; $P = 0.007$). Performance estimations indicated that this score added beyond standard clinical parameters and beyond both the *MGMT* methylation status and the *IDH1* mutational status in terms of discrimination (C statistics, 0.827 versus 0.835; $P < 0.001$).

Conclusion: The 4-gene signature provides an independent risk score strongly associated with outcome of patients with high-grade gliomas. *Clin Cancer Res; 17(2); 317–27. ©2011 AACR.*

Introduction

High-grade gliomas (HGG) are brain tumors associated with high morbidity and mortality. They are classified as either grade III or grade IV on the basis of histopathologic features established by the World Health Organization (WHO) (1). In combination with other clinical parameters, the grade has long provided important prognostic information (2). Recently, molecular biomarkers have been shown to be strongly associated with the prognostic of these tumors. O(6)-methylguanine-DNA-methyltransferase (*MGMT*) promoter hypo-methylation is involved in glioblastoma resistance to temozolomide chemotherapy (3) and mutations of the isocitrate dehydrogenase 1 (*IDH1*) gene are associated with better outcome of patients (4).

Recent studies have demonstrated that molecular and genetic analysis of gliomas could help in their classification and in the design of treatment protocols (5, 6). Microarray expression profiling has characterized molecular subtypes of brain tumors associated with tumor grade, progression, and prognosis (6–11), though only a few genes have been consistently identified (12). To overcome such a lack of reproducibility, the best approach is to analyze multiple data set simultaneously to combine the results from relevant studies. Such analysis applied to microarray data has been shown to be a powerful tool to identify candidate biomarkers and biological pathways (13).

The 2 most comprehensive glioma microarray classifications schemes published to date (6, 9) are based on unsupervised analysis, and they clearly show a strong association between the tumor grading and the defined

Authors' Affiliations: [1]CNRS UMR 6061 Genetic and Development, University of Rennes 1, Rennes, France; [2]Medical Genomics Unit, Molecular Genetics and Genomics, University Hospital Rennes, France; [3]Biogenouest Transcriptome Platform, University of Rennes 1, Rennes, France; [4]Departments of Pathology and [5]Neurosurgery, University Hospital Rennes, France; [6]Clinical Research Unit, Department of Medical Oncology and [7]Department of Clinical Biology, Eugène Marquis Cancer Institute, Rennes, France; and [8]Department of Neurosurgery, University Hospital Angers, France

Note: Supplementary data for this article are available at Clinical Cancer Research Online (http://clincancerres.aacrjournals.org/).

Marie de Tayrac and Marc Aubry contributed equally to the work.

Corresponding Author: Jean Mosser, CNRS UMR 6061 Genetic and Development, University of Rennes 1, 2, av du Pr Léon Bernard, Rennes 35000, France. Phone: 33 2 23 23 44 91, Fax: +33 2 23 23 44 78; E-mail: jean.mosser@univ-rennes1.fr.

doi: 10.1158/1078-0432.CCR-10-1126

Translational Relevance

In this study we report the development and validation of a risk-score model based on the weighted expression of 4 genes: *CHAF1B*, *EDNRB*, *HJURP*, and *PDLIM4*. This 4-gene risk score is highly associated with the outcome of patients with newly diagnosed high-grade gliomas independently from clinical risk factors, *IDH1* mutational status, and *MGMT* methylation status. These results suggest the importance of this multimarker panel as a stratification factor for the design of future comparative therapeutic trials.

glioma subtypes. These 2 classifications proposed by Phillips et al. (8) and Li et al. (6) show that HGG patients with better-than-expected survival could be classified in an enriched grade III subtype designated *proneural* or *oligodendroglioma-rich*, respectively.

Here we considered a supervised approach to account for the current WHO grading when deriving gene biomarkers associated with clinical outcomes (7, 10, 11). Following this rationale, we performed a combined analysis to identify such biomarkers from a robust signature related to tumor aggressiveness. We identified and validated a risk-score model based on the weighted expression of 4 genes: *CHAF1B*, *PDLIM4*, *EDNRB*, and *HJURP*. Two independent validations were performed: the first, on a public microarray study, and the second, on a new set of HGGs by quantitative reverse transcription-PCR (RT-qPCR). We also compared the performances of our risk-score model with the prognostic value of currently admitted clinical and molecular risk factors (*MGMT* promoter methylation status and *IDH1* mutational status).

Materials and Methods

Study samples

The local cohort totalized 194 patients with newly diagnosed and untreated HGGs admitted to the University hospitals involved in the French Cancéropôle Grand-Ouest Glioma Project (see Supplementary References). Patients were selected retrospectively in the period 1998 to 2008 with a follow-up time of minimum 2 years. Tumor samples were collected in accordance with the French regulations and the Declaration of Helsinki. All patients gave their informed consent before inclusion. Initial histology was confirmed by a central review involving at least 2 neuropathologists according to the WHO classification of central nervous system tumors (1). Patient characteristics are summarized in Table 1. Total DNAs and RNAs were isolated from frozen samples of primary brain tumors stored (−80°C) in the Cancéropôle Biological Resource Centers (AC-2008-77). Quality of DNA samples was assessed on 1% agarose gel and RNA integrity was confirmed using the Agilent 2100 Bioanalyzer (Agilent Technologies).

RT-qPCR analysis. RT-qPCR reactions were performed as described previously (14) with *B2M* (β-2 microglobulin) and *HPRT1* (hypoxanthine phosphoribosyltransferase) as internal controls. All the primer sequences used in this study can be found in the Supplementary Experimental Procedures.

***IDH1* mutations.** Exon 4 of the *IDH1* gene was amplified with the use of a PCR assay and sequenced in DNA from the tumor from each patient, as described previously (15). Patients were screened for somatic mutations affecting the R132 residue of *IDH1*.

***MGMT* promoter methylation.** The pyrosequencing methylation assay was performed with the PyroMark Q96 CpG MGMT kit (Qiagen), according to the manufacturer's protocol. Samples were considered methylated if they had average CpG methylation of 9% or more and unmethylated if they had average methylation less than 9%, in duplicate reactions (16).

External data collection

External microarray data for 326 patients were collected from 4 Gene Expression Omnibus HGGs data sets (7, 8, 10, 17). There were 22,215 common probe sets in the 3 data sets. Base-10 log-transformed intensities were centered using the scale function of the R base package. Data sets characteristics and analysis workflow are presented in Fig. 1.

Statistical analysis

Combined analysis of microarray data. Combined analysis was performed on 267 patients (GDS1962, GSE4271, and GSE4412) using the Bioconductor *RankProd* package (18). This package utilizes the rank product nonparametric method to identify up- and downregulated genes between anaplastic astrocytomas and glioblastomas (19). The *RankProd* package was chosen for its ability to easily combine data sets from different origins (laboratories and environments) into a single analysis. It was also shown that this nonparametric method outperforms other meta-analysis methods in terms of sensitivity and specificity (18). Individual analyses were also performed for each study (2-sided Student's *t* test) and results were combined. Genes were considered to be differentially expressed for a corrected *P* value (false discovery rate, FDR) less than 0.05 and a fold change greater than 2 in at least 1 of the 2 approaches. Functional annotation analyses were assessed using the Database for Annotation, Visualisation, and Integrated Discovery (http://david.abcc.ncifcrf.gov/) and unsupervised PCA with integration of biological knowledge (20). We used Benjamini corrected *P* values for multiple testing ($P < 0.05$).

Survival analysis and prognostic model selection. We performed a cross-study analysis of genes that can assist in the prognostication of survival by univariate Cox regression analysis. Gene expression was used as a predictor and survival time (in months) as the response. To select the significant genes, we controlled the FDR with the Benjamini–Hochberg correction and set the *P* value threshold at 0.01. To build an optimal survival model, we selected survival-associated genes with the *rbsurv* R package. Briefly, this package allows a sequential selection of genes based on the Cox proportional hazard model and on maximization of log-likelihood. To increase robustness, this package also

Table 1. Patients' clinical characteristics and stratification on the 4-gene expression risk score

Characteristic	All patients ($N = 194$)	Patients with low risk score ($N = 55$)	Patients with high risk score ($N = 139$)
Age, y			
Median	57	52	58
Range	13–80	13–77	16–80
Age, *n* (%)			
≤50 y	64 (33)	25 (46)	39 (28)
>50 y	130 (67)	30 (54)	100 (72)
Univariate analysis	$P = 0.006$		
Sex, *n* (%)			
Male	103 (53)	32 (58)	71 (51)
Female	91 (47)	23 (42)	68 (49)
Univariate analysis	$P = 0.85$		
Preoperative KPS performance status (%)			
Median	80	85	80
Range	20–100	40–100	20–100
ND, *n*	15	9	6
Univariate analysis	$P = 0.692$		
Extent of surgery, *n* (%)			
None	2 (1)	0 (0)	2 (1)
Biopsy	13 (7)	5 (9)	8 (6)
Debulking			
Partial resection	49 (25)	15 (28)	34 (25)
Complete resection	123 (63)	34 (62)	89 (64)
ND	7 (4)	1 (1)	6 (4)
Univariate analysis	$P = 0.438$		
RTOG RPA classification, *n* (%)			
I–II	26 (14)	18 (33)	8 (6)
III–IV	66 (34)	17 (31)	49 (35)
V–VI	99 (51)	19 (35)	80 (58)
ND	3 (2)	1 (1)	2 (1)
Univariate analysis	$P < 0.001$		
Therapy, *n* (%)			
None	4 (2)	1 (1)	3 (2)
Radiotherapy alone	20 (10)	5 (9)	15 (11)
Chemotherapy alone	7 (4)	5 (9)	2 (1)
Radiotherapy plus chemotherapy			
Temozolomide	106	18 (33)	88 (63)
PCV[a]	28	19 (35)	9 (7)
Other[b]	27 (14)	5 (9)	22 (16)
ND	2 (1)	2 (4)	0
Univariate analysis	$P = 0.366$		
IDH1 mutation, *n* (%)			
Mutated[c]	30 (15)	20 (37)	10 (7)
Wild-type	159 (82)	32 (58)	127 (92)
ND	5 (3)	3 (5)	2 (1)
Univariate analysis	$P < 0.001$		
MGMT status, *n* (%)			
Unmethylated	94 (49)	20 (37)	74 (53)
Methylated	90 (46)	32 (58)	58 (42)
ND	10 (5)	3 (5)	7 (5)
Univariate analysis	$P < 0.001$		

(Continued on the following page)

Table 1. Patients' clinical characteristics and stratification on the 4-gene expression risk score (cont'd)

Characteristic	All patients ($N = 194$)	Patients with low risk score ($N = 55$)	Patients with high risk score ($N = 139$)
Findings on pathologic review, *n*			
Glioblastoma[d]	145	23	122
Anaplastic astrocytoma[e]	38	22	16
With necrosis and vascular proliferation	25	13	12
Anaplastic oligodendroglioma	11	10	1
With necrosis and vascular proliferation	3	2	1
Univariate analysis	$P < 0.001$		
Survival, mo			
Median	16.2	55.8	14.5
95% CI	14.7–18.3	26.0 to NR	12.5–16.0

[a]PCV consists of 3 chemotherapy drugs: Procarbazine, CCNU, and Vincristine.
[b]Other: includes topotecan, BCNU, Gemini, and "8 drugs in one EORTC trial" chemotherapy.
[c]See Supplementary Table S5 for details about sequencing results.
[d]Glioblastoma included 4 secondary glioblastomas.
[e]Anaplastic astrocytoma included oligoastrocytoma.
Abbreviation: NR, median survival not reached.

selects survival-associated genes by repetition (1,000 times) of a separation between the training and validation sets of samples. Regression coefficients of the optimal survival model were estimated after adjustment on the study factor. Risk scores were determined using classical Cox model risk formulae with a linear combination of the gene expression values weighted by the estimated regression coefficients. Time-dependent receiver-operating characteristic (ROC) curve analyses were used to select the optimal risk cutoffs for the stratification of patients. The Kaplan–Meier method was used to estimate the survival distributions. Log-rank tests were used to test the difference between survival groups. Analyses were carried out with the *survival* and *survivalROC* R packages.

Prognostic model validation and performances. We constructed a model including clinical factors—age, treatment, histologic grade, and risk classes as defined by the Radiation Therapy Oncology Group (RTOG) by recursive partitioning analysis (RPA; ref. 21)—along with *MGMT* methylation status and *IDH1* mutational status. We evaluated the discriminatory capability of the model with the gene expression risk score as compared with the model without the gene expression risk score using C statistics. Differences in discrimination were evaluated using a nonparametric approach (22). We assessed model calibration using the Hosmer–Lemeshow χ^2 test (23). Analyses were performed using the *Hmisc* and *Design* R packages.

Results

Data sets characteristics and analysis workflow are summarized in Fig. 1.

Consensus gene selection in high-grade gliomas

Combined analysis and individual study approaches were performed to define a consensus gene expression signature in HGGs that could be used to find biomarkers associated with clinical outcomes. This signature was composed of 438 gene probe sets with 65 identified by both approaches (Supplementary Fig. S2; Supplementary Table S1). Associated enriched GO processes (Supplementary Tables S2 and S3) were related to invasion (Supplementary Fig. S5), angiogenesis (Supplementary Fig. S3), response to stress (Supplementary Fig. S6), and morphogenesis (Supplementary Fig. S4). Among the consensus genes strongly associated with grading, we selected and validated 9 genes (*CHI3L1*, *ADAM12*, *S100A4*, *TIMP1*, *NDRG2*, *NTRS2*, *LUZP2*, *ALDH5A1*, and *RASL10A*) by RT-qPCR ($P < 0.001$) on a subset of 90 HGG samples (Supplementary Fig. S7).

A gene expression risk score associated with survival in high-grade gliomas

To assess the survival prognosis capabilities of the 438 selected probe sets, we performed univariate Cox analyses of the expression data for these genes, with overall survival (OS) as a dependent variable. We ranked the genes on the basis of their predictive power (univariate *z* score). We then selected the genes having a highly significant association with survival and identified 40 genes with high predictive power (Supplementary Fig. S8). According to the univariate *z* score, 26 were risk genes and 14 were protective genes. Risk genes were related to gene ontology (GO) biological process *cell cycle* (*CDC25A*, *ASPM*, *CHAF1B*, *CENPE*, *CEP55*, *CDC20*, *NCAPG*, *AURKA*) and to *ECM-receptor interaction* and *Focal Adhesion* KEGG pathways (*HMMR*, *COL1A2*, *COL4A2*,

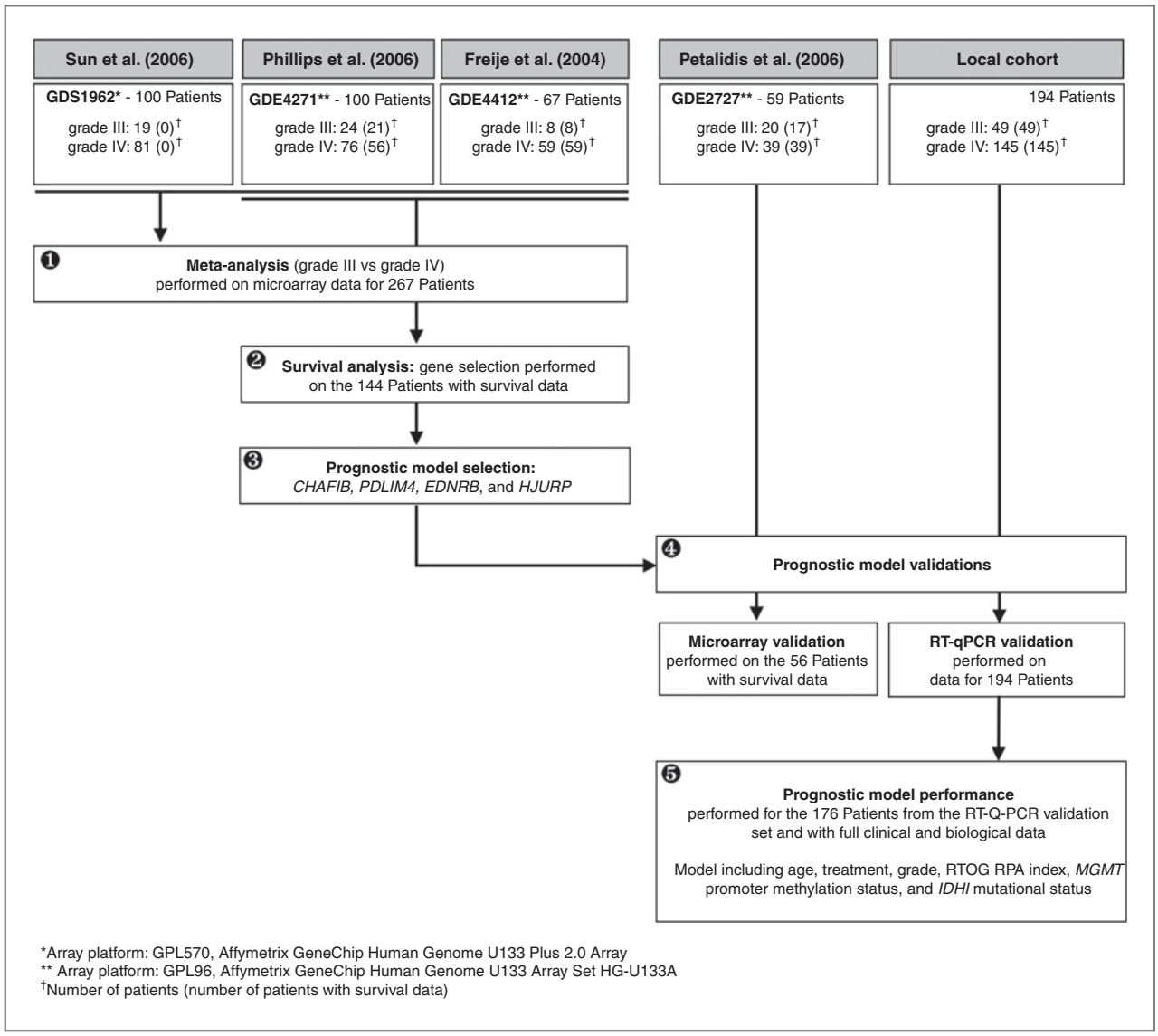

Figure 1. Analysis workflow. (1) Meta-analysis was performed on 3 publicly available HGGs microarray data sets (267 patients) to define a robust signature related to tumor aggressiveness (grade III versus grade IV). (2) This signature was used to define genes also associated with outcome by survival analysis. This was performed on 144 of the 267 patients for which survival data was available. (3) Genes associated with both grading and outcome were used to select an optimal survival model. This model was based on the weighted expression of 4 genes (risk score). (4) Two independent validations were performed: the first, on a publicly available microarray study, and the second, on the local cohort of HGGs, by RT-qPCR. (5) Model performances were assessed on the patients from the local cohort with full clinical and biological data (176 of 194 patients).

COL1A1, *COL4A1*, *MET*). Interestingly, 5 of the protective genes were related to GO biological process *nervous system development* (*EDNRB*, *ABLIM1*, *ALDH5A1*, *NDRG2*, *FGF12*).

We performed multivariate Cox regression analyses to create an optimal gene-based survival model. We used the 40 selected genes to sequentially construct survival models. The model best associated with survival ($P < 0.001$) and with good discrimination ability (C statistic, 0.843; 95% CI, 0.647–0.827) was based on the expression of 4 genes: *CHAF1B*, *PDLIM4*, *EDNRB*, and *HJURP*. The relative contributions of each of the 4 genes in the multivariate analysis are summarized in the portion of the Cox risk equation that captures the individual risk profile: $(0.587 \times CHAF1B) + (0.326 \times PDLIM4) + (-0.470 \times EDNRB) + (0.532 \times HJURP)$. Patients were ranked according to their risk score. The optimal risk cutoff was assessed and used for the stratification of patients into 2 groups: low risk of death and high risk of death. Patients with a low risk of death (25 anaplastic astrocytomas and 36 glioblastomas) had a median OS of 46.6 months (95% CI, 28.7–73.9), which was significantly longer than 11.7 months (95% CI, 9.0–13.5) for patients with a high risk of death (4 anaplastic astrocytomas and 79 glioblastomas), $P < 0.001$ by the log-rank test (Fig. 2A).

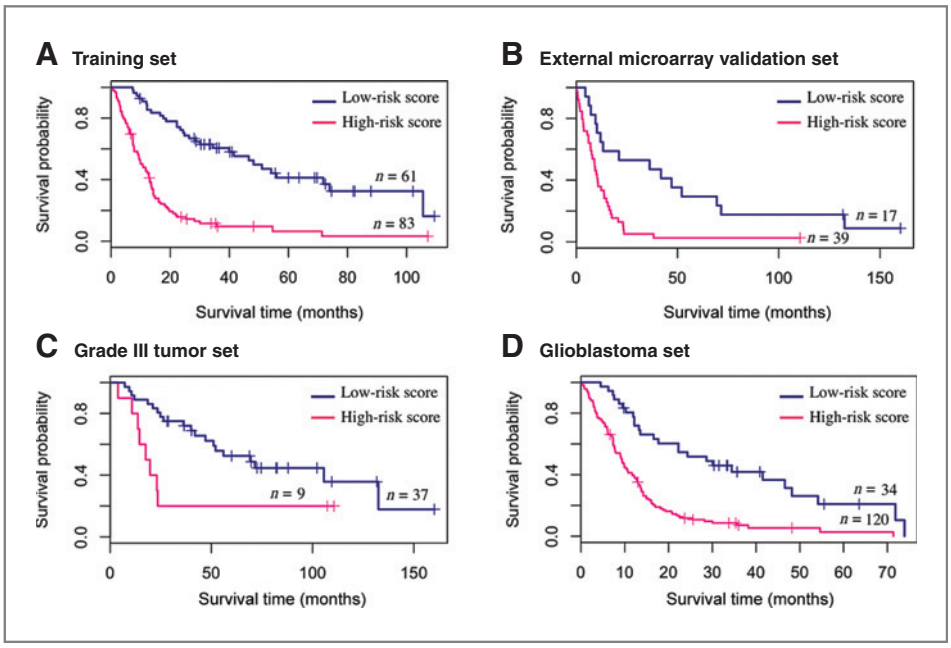

Figure 2. Kaplan–Meier estimates of overall survival after subdivision into low and high risk-score groups. A, 144 patients with malignant glioma, analyzed by microarray meta-analysis (8, 10). B, 56 patients with malignant gliomas reported by Petalidis et al. (7). C, whole anaplastic astrocytoma set ($n = 46$). D, whole glioblastoma set ($n = 154$).

During this work, the MD Anderson group published a 9-gene panel (*AQP1*, *CHI3L1*, *EMP3*, *GPNMB*, *IGFBP2*, *LGALS3*, *OLIG2*, *PDPN*, and *RTN1*) to predict outcome in glioblastoma (24). Six of these genes were also found in our consensus gene selection. We compared our 4-gene panel with the MD Anderson group 9-gene predictor. Both models were highly significant ($P = 1e\text{-}08$ and $P = 3e\text{-}05$, respectively). The discrimination of the 4-gene model was significantly higher than the discrimination of the 9-gene model [C statistic, 0.80 (95% CI, 0.72–0.86) versus 0.76 (95% CI, 0.64–0.81), $P < 0.001$, respectively; Supplementary Fig. S9], showing the relevance of the 4-gene panel.

We performed an external validation of the 4-gene survival model with an independent microarray data set comprising 56 HGGs with survival data reported by Petalidis et al. (7). Patients were divided into 2 groups on the basis of the 4-gene model (low or high risk of death). The low-risk group was composed of 12 anaplastic astrocytomas and 5 glioblastomas and the high-risk group of 5 anaplastic astrocytomas and 34 glioblastomas. The OS was higher in low-risk HGGs than in high-risk HGGs [17.8 months (95% CI, 9.6–47.9) versus 9.3 months (95% CI, 7.2–13.9), respectively; $P < 0.001$; Fig. 2B]. The discrimination was as good as in the original data (C statistic, 0.852; 95% CI, 0.673–0.933).

Model validation was also performed to determine whether the 4-gene expression data contained survival-predictive information that was distinct from the prediction embedded within histologic grade. In the whole anaplastic astrocytoma set, the OS was higher in low-risk patients ($n = 9$) than in high-risk patients [$n = 37$; 69.4 months (95% CI, 41.8 to not reached) versus 19.7 months (95% CI, 13.7 to not reached), respectively; $P < 0.05$; Fig. 2C]. In the whole glioblastoma set, low-risk patients ($n = 34$) had a much higher OS (30.07 months; 95% CI, 17.7–54.2) than high-risk patients ($n = 120$; 9.3 months; 95% CI, 7.6–11.7; $P < 0.001$; Fig. 2D).

Evaluation of the gene expression risk-score performances

A cohort of 194 patients with extensive bioclinical parameters was used to validate the performances of the 4-gene classifier (Table 1). Univariate analyses showed that the gene expression risk score, the DNA methylation status of the *MGMT* promoter, and the *IDH1* mutational status were significantly associated with the OS in this cohort. In the whole cohort, patients were divided into 2 groups on the basis of the risk-score model with log2-transformed data issued from RT-qPCR analysis. The OS was clearly higher for low-risk patients (55.8 months; 95% CI, 26.0 to not reached) than for high-risk patients (14.5 months; 95% CI, 12.5–16.0; $P < 0.001$; Fig. 3A). In this population, *MGMT*-methylated tumors, compared with unmethylated tumors, had a significantly better OS [19.5 months (95% CI, 16.7–29.4) versus 14.5 months (95% CI, 11.4–16.2), respectively; $P < 0.001$; Fig. 3B]. Similarly, in this group, *IDH1*-mutated tumors had a much higher OS (median survival not reached; 95% CI, 42.5 to not reached) than *IDH1*-nonmutated tumors (14.9 months; 95% CI, 13.7–16.5; $P < 0.001$; Fig. 3C).

Two multivariate models were built, both including age, treatment, grade, RTOG RPA classes, *MGMT* methylation status, and *IDH1* mutational status; 1 with and 1 without the 4-gene expression risk score. These models were used to estimate the prognostic value of the gene expression risk score (i) for the 176 of 194 patients with complete data for all variables and (ii) for a subset of patients treated with temozolomide chemoradiation ($n = 105$). Results are

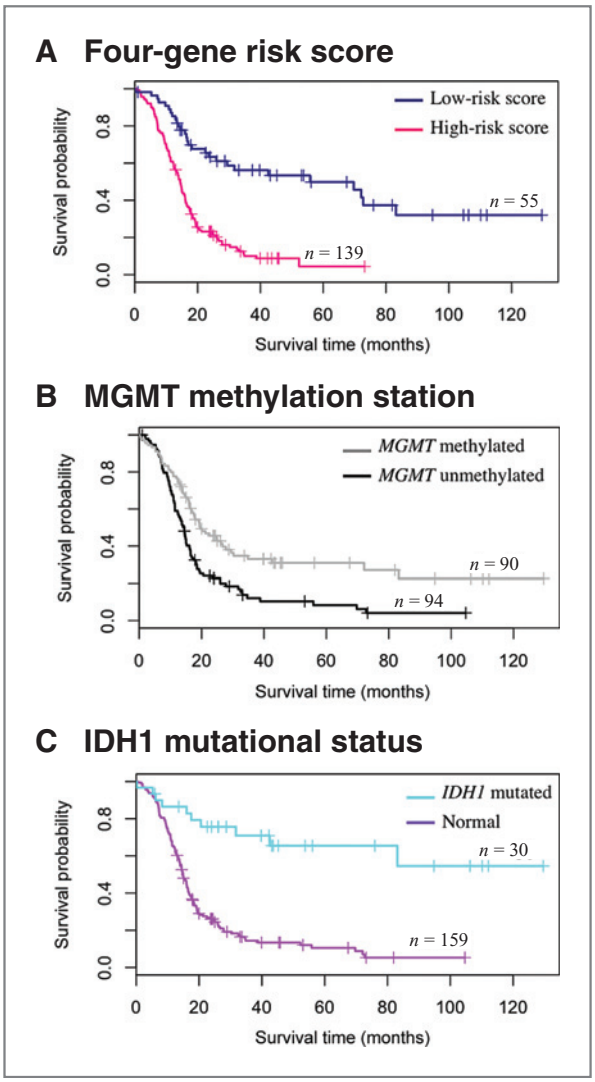

Figure 3. Survival of patients with high-grade glioma according to the 4-gene risk score, the *MGMT* promoter methylation status and the *IDH1* mutational status. A, Kaplan–Meier estimates of overall survival in the whole local cohort after subdivision into 2 groups (low and high risk of death) on the basis of the risk-score model, with log2-transformed data issued from quantitative RT-qPCR analysis. The overall survival among low-risk patients is 55.8 months (95% CI, 26.0 to not reached), as compared with 14.5 months (95% CI, 12.5–16.0) among high-risk patients ($P < 0.001$). B, Kaplan–Meier estimates of overall survival in the whole local cohort after subdivision into 2 groups depending on the DNA methylation status of the *MGMT* promoter. Median survival is 19.5 months (95% CI, 16.7–29.4) for patients with tumoral methylated *MGMT* promoter and 14.5 months (95% CI, 11.4–16.2) for patients with tumoral unmethylated *MGMT* promoter. C, Kaplan–Meier estimates of overall survival in the whole local cohort after subdivision into 2 groups depending on the presence of *IDH1* mutations. *IDH1* mutational status is significantly associated with the overall survival in all cohorts [$P < 0.001$, median survival not reached (95% CI, 42.5 to not reached) versus 14.9 months (95% CI, 13.7–16.5)].

provided in Table 2. In both cases, the gene expression risk score was strongly associated with survival (hazard ratio = 0.49; 95% CI, 0.30–0.81; $P = 0.005$; and hazard ratio = 0.37; 95% CI, 0.18–0.77; $P = 0.008$; respectively) and all models showed excellent discrimination, with C statistics more than 0.80. In the whole cohort and for the patients treated with temozolomide chemotherapy, the C statistic improved significantly with the addition of the gene expression risk score in the model (0.816 versus 0.846, $P < 0.001$ and 0.792 versus 0.822, $P < 0.001$, respectively), showing that the 4-gene risk score added beyond standard clinical parameters and beyond both the *MGMT* methylation status and the *IDH1* mutational status.

The performances of the gene expression risk score were also evaluated on a subset of 98 patients with glioblastoma who underwent tumor resection and who were treated with radiotherapy plus concomitant and adjuvant temozolomide. After adjustment for RTOG RPA classes and *MGMT* promoter methylation status, multivariate analysis confirmed that the 4-gene expression risk score was an independent marker robustly associated with outcome for glioblastoma patients treated with standard protocol (hazard ratio = 0.386; 95% CI, 0.164–0.910; $P = 0.03$).

Discussion

Molecular studies of HGGs have highlighted the heterogeneity of these tumors, and have linked molecular signatures to their natural history and to differences in survival rates. It is likely that the ability to identify such molecular subtypes of tumors will be essential for guiding therapeutic advances. In this study we report a risk-score model based on the expression of 4 genes for the stratification of patients with HGGs. This risk calculation is based on a consensus gene expression signature and is strongly associated with survival independently from current clinical risk factors, *IDH1* mutational status, and *MGMT* promoter methylation status.

The initial step of our study consisted in a discovery phase for the identification of biomarkers repeatedly correlated with both tumor aggressiveness and patient outcome. It should be noticed that information regarding the therapeutic regimens was not incorporated in the meta-analysis of microarray data sets. Although this could have weakened this discovery phase, combining multiple and independent data sets was also an asset to identify robust biomarkers. Moreover, the RT-qPCR validation of the 4-gene signature in an external cohort of patients showed that the 2 risk groups had significant differences in OS independently from treatment. These results suggest that the 4 genes are relevant molecular markers in HGGs.

One explanation for the association between the 4-gene signature and clinical outcome could be that it may detect the molecular fingerprints inherent to glioma aggressiveness. The proposed multimarker panel is based on the expression of *EDNRB*, *CHAF1B*, *PDLIM4*, and *HJURP*. In this model, the overexpression of *EDNRB* correlates with better prognosis. *EDNRB* encodes the endothelin receptor type B implicated in tumor proliferation, survival, invasion, angiogenesis, and metastasis (25). Freije et al. (10) have reported *EDNRB* as a member of the neurogenesis-related genes group that portends the longest survival.

Table 2. Comparison of prognostic model adjusted for clinical factors along with MGMT promoter methylation status and IDH1 mutational status, with or without the 4-gene risk score

	Prediction Model	
	Without the 4-gene expression risk score	**With the 4-gene expression risk score**
Whole cohort (n = 176)		
Age <50 y vs >50 y		
Hazard ratio (95% CI)	0.99 (0.97–1.01)	0.99 (0.97–1.01)
P	0.47	0.56
RTOG RPA classification, per unit increase		
Hazard ratio (95% CI)	1.05 (0.71–1.59)	1.02 (0.68–1.53)
P	0.78	0.93
Treatment, per unit increase		
Hazard ratio (95% CI)	0.81 (0.66–0.98)	0.83 (0.69–1.01)
P	0.03	0.07
Histology, grade IV vs III		
Hazard ratio (95% CI)	3.28 (1.74–6.14)	1.62 (0.84–3.13)
P	<0.001	0.01
MGMT methylated vs unmethylated		
Hazard ratio (95% CI)	0.61 (0.43–0.87)	0.61 (0.42–0.88)
P	0.007	0.007
IDH1 mutated vs nonmutated		
Hazard ratio (95% CI)	0.32 (0.14–0.71)	0.38 (0.17–0.84)
P	0.005	0.02
Four-gene risk score, low vs high		
Hazard ratio (95% CI)	-	0.49 (0.30–0.81)
P	-	0.005
Discriminatory capability		
C statistic (95% CI)	0.816 (0.739–0.891)	0.846 (0.770–0.913)
P value for difference	<0.001	
Accuracy of calibration at 3 y		
χ^2	3.61	3.57
P	0.935	0.937
Patients treated with temozolomide chemoradiation (n = 105)		
Age <50 y vs >50 y		
Hazard ratio (95% CI)	1.00 (0.97–1.03)	1.00 (0.97–1.03)
P	0.97	0.98
RTOG RPA classification, per unit increase		
Hazard ratio (95% CI)	1.22 (0.58–2.61)	1.34 (0.66–2.80)
P	0.59	0.43
Histology, grade IV vs III		
Hazard ratio (95% CI)	1.67 (0.49–5.60)	1.06 (0.30–3.75)
P	0.41	0.92
MGMT methylated vs unmethylated		
Hazard ratio (95% CI)	0.60 (0.37–0.95)	0.53 (0.33–0.86)
P	0.03	0.01
IDH1 mutated vs nonmutated		
Hazard ratio (95% CI)	0.10 (0.01–0.77)	0.11 (0.01–0.89)
P	0.03	0.04
Four-gene risk score, low vs high		
Hazard ratio (95% CI)	-	0.37 (0.18–0.78)
P	-	0.008

(Continued on the following page)

Table 2. Comparison of prognostic model adjusted for clinical factors along with MGMT promoter methylation status and IDH1 mutational status, with or without the 4-gene risk score (cont'd)

	Prediction Model	
	Without the 4-gene expression risk score	**With the 4-gene expression risk score**
Discriminatory capability		
C statistic (95% CI)	0.793 (0.592–0.937)	0.821 (0.688–0.903)
P value for difference	<0.001	
Accuracy of calibration at 3 y		
χ^2	3.55	3.58
P	0.939	0.937

The 3 other genes of our model (*CHAF1B*, *PDLIM4*, *HJURP*) are correlated with a higher risk of death. *CHAF1B* encodes the p60 subunit of the chromatin assembly factor I, which plays a major role in chromatin assembly after replication and DNA repair. It has been proposed as a specific marker of actively proliferating cells (26) and as a predictor of poor outcome in squamous cell carcinoma of the tongue (27). *PDLIM4*, a LIM domain gene also known as RIL, is suspected to have tumor suppressor functions in myeloid diseases (28) and prostate cancer (29) by loss of heterozygosity (LOH), deletion, or hypermethylation. However, its extreme upregulation by integrin-promoted demethylation has been recently reported (30) in breast carcinoma cells together with other genes also validated in our study (*S100A4*, *NCAPG*), suggesting a potential oncogenic function of *PDLIM4*. The Holliday Junction Recognition Protein (*HJURP*) was recently shown to be an indispensable factor for cell-cycle regulation of centromeric chromatin assembly (31, 32) and for chromosomal stability in immortalized cancer cells (33). It has also recently been suggested that *HJURP* could be implicated in glioma malignancy (34). These studies and our findings suggest that these 4 genes are important molecular components of astrocytic tumors aggressiveness.

The 2 risk groups defined by the 4-gene classifier are also characterized by the expression change of genes related to cancer malignancy or survival of gliomas (Supplementary Table S4). Genes highly expressed in high-risk HGGs are remarkably related to cell cycle and cytokinesis, in accordance with the fact that aggressive tumors exhibit a high percentage of cycling cells. This was also reported for the *Proliferative* subgroup of HGGs identified by Philips et al. (8). Most of the genes highly expressed in low-risk HGGs are related to the development of the nervous system. Other authors (8, 10, 35) also described a correlation between neuronal markers and the favorable subclasses of HGGs. These findings underline that the 2 risk groups have distinct molecular phenotypes and suggest that they may respond differently to therapeutic regimens.

Multivariate analysis confirmed that both the mutations of *IDH1* and the presence of *MGMT* promoter methylation were associated with a survival benefit in the whole cohort of HGGs and in the subgroup of patients with glioblastoma treated similarly with temozolomide chemoradiation. This analysis also showed that the 4-gene expression risk score was strongly associated with outcome, independently from clinical and molecular risk factors. The performance evaluation indicated that the 4 genes added beyond the prognostic capabilities of all these factors. These results suggest that the 4-gene status, along with the existing clinical and other molecular markers, could be used to optimize patient stratification. As an illustration, when combined with the *IDH1* mutational status, the 4-gene risk score allowed the identification of 3 groups of HGGs (good-, intermediate-, and poor-outcome groups) with significant differences in OS ($P < 0.001$; Fig. 4). The group of HGGs with intermediate-outcome (nonmutated/low-risk or

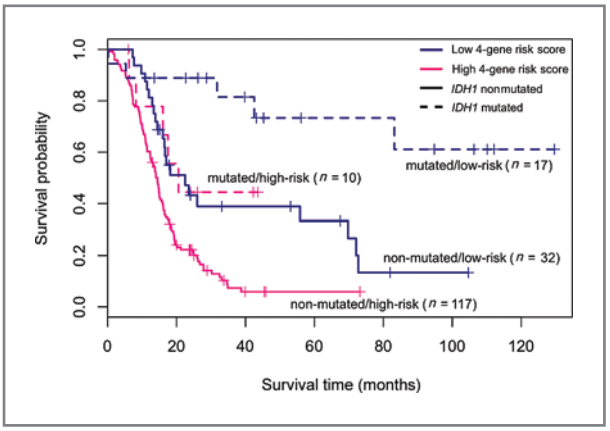

Figure 4. Combined stratification based on the *IDH1* mutational status and the 4-gene risk score. Three groups of HGGs (good-, intermediate-, and poor-outcome groups) with significant differences in OS ($P < 0.001$) are defined by the combination of the *IDH1* mutational status and the 4-gene risk score. The group of HGGs with intermediate-outcome (nonmutated/low-risk or mutated/high-risk) is characterized by a median survival of 20.6 months (95% CI, 16.5–72.1), as compared with 14 months (95% CI, 12.3–15.2) for the poor-outcome group (nonmutated/high-risk) and to a median survival not reached (95% CI, 83.2 to not reached) for the good-outcome group (mutated/low-risk).

mutated/high-risk) was characterized by a median survival of 20.6 months (95% CI, 16.5–72.1), as compared with 14 months (95% CI, 12.3–15.2) for the poor-outcome group (nonmutated/high-risk) and with a median survival not reached (95% CI, 83.2 to not reached) for the good-outcome group (mutated/low-risk). For this intermediate-outcome group (representing 24% of the whole cohort), the *MGMT* methylation status did not provide any predictive information ($P = 0.5$) and the median survival time was similar to that of patients with methylated *MGMT* promoter. These results suggest the importance of taking into account the 4-gene signature as a stratification factor for the design of future comparative therapeutic trials, though it needs to be further investigated in a prospective clinical study.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Grant Support

This work was supported by grants from the Fonds européen de développement régional (FEDER), the Cancéropôle Grand-Ouest Glioma Project, Institut national du cancer (INCa), Ligue contre le Cancer foundation (LNCC), Région Bretagne (PRIR), and CRITT Santé Bretagne.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received April 29, 2010; revised July 8, 2010; accepted August 17, 2010; published OnlineFirst January 11, 2011.

References

1. Louis DN, Ohgaki H, Wiestler OD, Cavenee WK, Burger PC, Jouvet A, et al. The 2007 WHO classification of tumours of the central nervous system. Acta Neuropathol 2007;114:97–109.
2. Louis DN. Molecular pathology of malignant gliomas. Annu Rev Pathol 2006;1:97–117.
3. Hegi ME, Diserens AC, Gorlia T, Hamou MF, de Tribolet N, Weller M, et al. MGMT gene silencing and benefit from temozolomide in glioblastoma. N Engl J Med 2005;352:997–1003.
4. Yan H, Parsons DW, Jin G, McLendon R, Rasheed BA, Yuan W, et al. IDH1 and IDH2 mutations in gliomas. N Engl J Med 360:765–73.
5. Behin A, Hoang-Xuan K, Carpentier AF and Delattre JY. Primary brain tumors in adults. Lancet 2003;361:323–31.
6. Li A, Walling J, Ahn S, Kotliarov Y, Su Q, Quezado M, et al. Unsupervised analysis of transcriptomic profiles reveals six glioma subtypes. Cancer Res 2009;69:2091–9.
7. Petalidis LP, Oulas A, Backlund M, Wayland MT, Liu L, Plant K, et al. Improved grading and survival prediction of human astrocytic brain tumors by artificial neural network analysis of gene expression microarray data. Mol Cancer Ther 2008;7:1013–24.
8. Phillips HS, Kharbanda S, Chen R, Forrest WF, Soriano RH, Wu TD, et al. Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis. Cancer Cell 2006;9:157–73.
9. Liang Y, Diehn M, Watson N, Bollen AW, Aldape KD, Nicholas MK, et al. Gene expression profiling reveals molecularly and clinically distinct subtypes of glioblastoma multiforme. Proc Natl Acad Sci USA 2005;102:5814–9.
10. Freije WA, Castro-Vargas FE, Fang Z, Horvath S, Cloughesy T, Liau LM, et al. Gene expression profiling of gliomas strongly predicts survival. Cancer Res 2004;64:6503–10.
11. Nutt CL, Mani DR, Betensky RA, Tamayo P, Cairncross JG, Ladd C, et al. Gene expression-based classification of malignant gliomas correlates better with survival than histological classification. Cancer Res 2003;63:1602–7.
12. Colman H, Aldape K. Molecular Predictors in Glioblastoma: toward Personalized Therapy. Arch Neurol 2008;65:877–83.
13. Hong F, Breitling R. A comparison of meta-analysis methods for detecting differentially expressed genes in microarray experiments. Bioinformatics 2008;24:374–82.
14. de Tayrac M, Etcheverry A, Aubry M, Saïkali S, Hamlat A, Quillien V, et al. Integrative genome-wide analysis reveals a robust genomic glioblastoma signature associated with copy number driving changes in gene expression. Genes Chromosomes Cancer 2009;48:55–68.
15. Parsons DW, Jones S, Zhang X, Lin JC, Leary RJ, Angenendt P, et al. An integrated genomic analysis of human glioblastoma multiforme. Science 2008;321:1807–12.
16. Dunn J, Baborie A, Alam F, Joyce K, Moxham M, Sibson R, et al. Extent of MGMT promoter methylation correlates with outcome in glioblastomas given temozolomide and radiotherapy. Br J Cancer 2009;101:124–31.
17. Sun L, Hui A-M, Su Q, Vortmeyer A, Kotliarov Y, Pastorino S, et al. Neuronal and glioma-derived stem cell factor induces angiogenesis within the brain. Cancer Cell 2006;9:287–300.
18. Hong F, Breitling R, McEntee CW, Wittner BS, Nemhauser JL and Chory J., et al. RankProd: a bioconductor package for detecting differentially expressed genes in meta-analysis. Bioinformatics 2006;22:2825–7.
19. Breitling R, Armengaud P, Amtmann A and Herzyk P. Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. FEBS Lett 2004; 573:83–92.
20. de Tayrac M, Lé S, Aubry M, Mosser J, and Husson F. Simultaneous analysis of distinct Omics data sets with integration of biological knowledge: Multiple Factor Analysis approach. BMC Genomics 2009;10:32.
21. Curran WJ Jr, Scott CB, Horton J, Nelson JS, Weinstein AS, Fischbach AJ, et al. Recursive partitioning analysis of prognostic factors in three Radiation Therapy Oncology Group malignant glioma trials. J Natl Cancer Inst 1993;85:704–10.
22. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988;44:837–45.
23. Hosmer DW Jr, Lemeshow S. Applied logistic regression. New York: John Wiley; 1989.
24. Colman H, Zhang L, Sulman EP, McDonald JM, Shooshtari NL, Rivera A, et al. A multigene predictor of outcome in glioblastoma. Neuro Oncol 2010;12:49–57.
25. Nelson J, Bagnato A, Battistini B, Nisen P. The endothelin axis: emerging role in cancer. Nat Rev Cancer 2003;3:110–6.
26. Polo SE, Theocharis SE, Klijanienko J, Savignoni A, Asselain B, Vielh P, et al. Chromatin assembly factor-1, a marker of clinical value to distinguish quiescent from proliferating cells. Cancer Res 2004; 64:2371–81.
27. Staibano S, Mignogna C, Muzio LL, Mascolo M, Salvatore G, Di Benedetto M, et al. Chromatin assembly factor-1 (CAF-1)-mediated regulation of cell proliferation and DNA repair: a link with the biological behaviour of squamous cell carcinoma of the tongue?Histopathology 2007;50:911–9.
28. Boumber YA, Kondo Y, Chen X, Shen L, Gharibyan V, Konishi K, et al. RIL, a LIM gene on 5q31, is silenced by methylation in cancer and sensitizes cancer cells to apoptosis. Cancer Res 2007; 67:1997–2005.
29. Vanaja DK, Ballman KV, Morlan BW, Cheville JC, Neumann RM, Lieber MM, et al. PDLIM4 repression by hypermethylation as a potential biomarker for prostate cancer. Clin Cancer Res 2006; 12:1128–36.

30. Chen M, Sinha M, Luxon BA, Bresnick AR and O'Connor KL. Integrin alpha6beta4 controls the expression of genes associated with cell motility, invasion, and metastasis, including S100A4/metastasin. J Biol Chem 2009; 284:1484–94.
31. Foltz DR, Jansen LE, Bailey AO, Yates JR 3rd, Bassett EA, Wood S, et al. Centromere-specific assembly of CENP-a nucleosomes is mediated by HJURP. Cell 2009;137:472–84.
32. Dunleavy EM, Roche D, Tagami H, Lacoste N, Ray-Gallet D, Nakamura Y, et al. HJURP is a cell-cycle-dependent maintenance and deposition factor of CENP-A at centromeres. Cell 2009; 137:485–97.
33. Kato T, Sato N, Hayama S, Yamabuki T, Ito T, Miyamoto M, et al. Activation of Holliday junction recognizing protein involved in the chromosomal stability and immortality of cancer cells. Cancer Res 2007;67:8544–53.
34. Valente V, Teixeira S, Neder L, Okamoto OK, Oba-Shinjo SM, Marie SK, et al. Selection of suitable housekeeping genes for expression analysis in glioblastoma using quantitative RT-PCR. BMC Mol Biol 2009;10:17.
35. Shirahata M, Oba S, Iwao-Koizumi K, Saito S, Ueno N, Oda M, et al. Using gene expression profiling to identify a prognostic molecular spectrum in gliomas. Cancer Sci 2009;100:165–7.