Title
Establishment of Prognostic Models for Astrocytic and Oligodendroglial Brain Tumors with Standardized Quantification of Marker Gene Expression and Clinical Variables

Journal
Biomarker Insights, 5(5)

ISSN
1177-2719

Authors
Zhou, Yi-Hong
Hess, Kenneth R
Raj, Vinay R
et al.

Publication Date
2010

DOI
10.4137/bmi.s6167

Biomarker Insights

ORIGINAL RESEARCH

Establishment of Prognostic Models for Astrocytic and Oligodendroglial Brain Tumors with Standardized Quantification of Marker Gene Expression and Clinical Variables

Yi-Hong Zhou[1], Kenneth R. Hess[2], Vinay R. Raj[3], Liping Yu[4], Longjian Liu[5], Alfred W.K. Yung[6] and Mark E. Linskey[1]

[1]Department of Neurological Surgery, The University of California, Irvine, CA 92697, USA. [2]Departments of Biostatistics, [6]Neuro-Oncology, The University of Texas M.D. Anderson Cancer Center, Houston, TX 77030, USA. [3]Division of Medical Genetics, The University of Arkansas for Medical Sciences, Little Rock, AR 72205, USA. [4]Ziren Research LLC, Irvine, CA 92618, USA. [5]Department of Epidemiology and Biostatistics, Drexel University School of Public Health, Philadelphia, PA 19102, USA. Corresponding author email: yihongz@uci.edu

Abstract

Background: Prognosis models established using multiple molecular markers in cancer along with clinical variables should enable prediction of natural disease progression and residual risk faced by patients. In this study, multivariate Cox proportional hazards analyses were done based on overall survival (OS) of 100 glioblastoma multiformes (GBMs, 92 events), 49 anaplastic astrocytomas (AAs, 33 events), 45 gliomas with oligodendroglial features, including anaplastic oligodendroglioma (AO, 13 events) and oligodendraglioma (O, 9 events). The modeling included two clinical variables (patient age and recurrence at the time of sample collection) and the expression variables of 13 genes selected based on their proven biological and/or prognosis functions in gliomas (*ABCG2*, *BMI1*, *MELK*, *MSI1*, *PROM1*, *CDK4*, *EGFR*, *MMP2*, *VEGFA*, *PAX6*, *PTEN*, *RPS9*, and *IGFBP2*). Gene expression data was a log-transformed ratio of marker and reference (*ACTB*) mRNA levels quantified using absolute real-time qRT-PCR.

Results: Age is positively associated with overall grade (4 for GBM, 3 for AA, 2_1 for AO_O), but lacks significant prognostic value in each grade. Recurrence is an unfavorable prognostic factor for AA, but lacks significant prognostic values for GBM and AO_O. Univariate models revealed opposing prognostic effects of *ABCG2*, *MELK*, *BMI1*, *PROM1*, *IGFBP2*, *PAX6*, *RPS9*, and *MSI1* expressions for astrocytic (GBM and AA) and oligodendroglial tumors (AO_O). Multivariate models revealed independent prognostic values for the expressions of *MSI1* (unfavorable) in GBM, *CDK4* (unfavorable) and *MMP2* (favorable) in AA, while *IGFBP2* and *MELK* (unfavorable) in AO_O. With all 13 genes and 2 clinical variables, the model R^2 was 14.2% ($P = 0.358$) for GBM, 45.2% ($P = 0.029$) for AA, and 62.2% ($P = 0.008$) for AO_O.

Conclusion: The study signifies the challenge in establishing a significant prognosis model for GBM. Our success in establishing prognosis models for AA and AO_O was largely based on identification of a set of genes with independent prognostic values and application of standardized gene expression quantification to allow formation of a large cohort in analysis.

Keywords: glioma, prognosis, model, gene expression markers

Biomarker Insights 2010:5 153–168

doi: 10.4137/BMI.S6167

This article is available from http://www.la-press.com.

Background

Glioma is the major part of primary malignant brain and central nervous system tumors characterized by tumor cell components of astrocytic glial, oligodendroglial or mixtures of both features. The most malignant and common form of brain tumor is glioblastoma multiforme (GBM, WHO grade IV), comprising about 50% of glioma. The one-year survival rate for GBM patients is 34% and two-year survival is only 12%.[1] Histopathological cellular morphology and tumor cytoarchitecture-based grading systems are used to classify gliomas, though providing useful prognostic diagnostic insights, these classification systems do not account for a significant proportion of variation in overall survival among individual glioma patients with the same histology, especially for non-GBM gliomas. Thus far, treatment options are limited for gliomas, for non-GBM gliomas in particular, to radiation and/or cytotoxic chemotherapy without stratification or triage. Prognosis models established using multiple molecular markers in cancer along with clinical variables should enable prediction of natural disease progression and residual risk faced by patients. For patients with glioma currently limited options exist in treatment with poor treatment efficacy. If we successfully achieve development and validation of prognostic models, we will have tools to enable patients to a starting point on the road to personalized treatment.

The power of a prognosis study relies on the number of events, and a meaningful analysis requires an average of 10 patient outcomes per variable used,[2] and clinic utility of a prognosis model requires standardization of data. Recent microarray-based gene expression profiling has provided molecular subclassifications of GBM and identification of genes with prognosis values.[3,4] However, direct application of a large panel of gene signature data generated by microarray based studies to modeling prognosis is challenged by the two main issues cited above. In an endeavor of applying gene expression information from tumors to explain patient survival variation, we have taken an alternate approach in modeling glioma prognosis. Our strategy is to establish a prognosis model based on a small to medium size of gene expression variables and a large size of samples and events.[5,6] The genes selected to be included in initial model establishment are based on their defined functions in cancer initiation and progression core pathways, and each prognosis value separately reported for glioma. In this study, we applied a standardized platform for real-time quantitative reverse transcription (qRT)-PCR to ensure data comparability.[7]

This study is a continuing exploration of modeling prognosis for gliomas with a select set of standardized gene expression data in a larger number of cDNA/tumor samples and patient's follow-up information together with two clinical variables (age and recurrence at the time to tumor sample collection) from the University of Texas, M.D. Anderson Cancer Center (MDACC), the University of Arkansas for Medical Science (UAMS), and the University of California Irvine (UCI). In addition to the prior studied cancer pathway related genes (*CDK4*, *EGFR*, *MMP2*, *VEGFA*, *PAX6*, *PTEN*, *RPS9*, and *IGFBP2*),[6] we included five cancer stem cell associated genes (*ABCG2*, *BMI1*, *MELK*, *MSI1*, *PROM1*) to explore their independent prognosis values in multivariate models for GBM (100 samples, 92 events), WHO grade III anaplastic astrocytoma (AA, 49 samples, 33 events), and oligodendroglial brain tumor, including a mix of WHO grade II oligodendroglioma (O, 18 samples, 9 events) and WHO grade III anaplastic oligodendroglioma (AO, 27 samples, 13 events).

The rationale for including the five stem cell associated genes in glioma prognosis modeling is on the emerging evidences on the existence of stem-like cells in brain tumors responsible for tumor resistance and recurrence.[8–11] The most well studied glioma stem-cell associated antigen is CD133[12] through expression of *PROM1* gene. The expression of other neural stem cell associated genes in glioma have also been reported to have adverse prognostic effects for patients, including *BMI1* in oligodendroglial tumors[13] and *MELK* in GBM of younger age patients.[14] Functionally involved in self-renewal of neural stem cells,[15] *BMI1* expression in glioma has been reported to determine tumor phenotype[16,17] and control chemo-response via activation of NF-kappaB signaling.[18] *MELK* expression has been shown to regulate the transition from GFAP-expressing progenitors to rapid amplifying progenitors in the postnatal brain[19] and promote glioma cell proliferation.[14]

The expression pattern and prognostic effects of two early identified stem cell associated genes, *ABCG2* and *MSI1*, have not been extensively studied

in gliomas. *ABCG2* is a marker of adult stem cells,[20] and *MSI1* marker of CNS stem cells and/or neural progenitor cells.[21,22] The quantities of (neural) stem cell associated gene expressions may represent the percentage of NSLC within the brain neoplasm and hence have prognostic values. In this study, we included all of the above five gene expression variables (*ABCG2*, *BMI1*, *MELK*, *MSI1*, and *PROM1*) in a multivariate model with two clinical (age and recurrence) and/or eight previously studied neoplastic pathway related genes (*CDK4*, *EGFR*, *MMP2*, *VEGFA*, *PAX6*, *PTEN*, *RPS9*, and *IGFBP2*)[6] to determine the overall contribution of each variable to glioma prognosis.

Methods

Patients and clinical data

Following informed consent, brain tumor specimens were collected from patients operated in M.D. Anderson Cancer Center (MDACC) at the University of Texas, the University of Arkansas for Medical Sciences (UAMS), and the University of California, Irvine (UCI) and included in this study. MDACC patients with AA, AO, and O were operated during 1987–1997. Majority of GBM patients (n = 59) of MDACC were operated during 1990–1997, while 7 in 2003. UAMS and UCI patients with GBM or AA, or AO were operated during 2003–2006. Following IRB approval, the clinical data for patient's age, recurrence (at time of sample collection), last contact date, and survival status were provided by each institutes tumor registry and verified by each site investigator in this study. The OS data was calculated based on the time of sample collection and time of death or last contact date.

Study design

This study was designed to establish three main prognosis models with standardized gene expression variables for three distinct gliomas based on histology: GBM model with 100 cases, AA model with 49 cases, and a model for glioma with oligodendroglial components: 27 anaplastic oligodendrogliomas (AO) and 12 oligodendrogliomas (O), as shown in Table 1. We combined AO and O into the same model (AO_O model) to make a total of 45 cases in order to have a meaningful analysis for this type of glioma. The patient survival data was mature for GBM (92% dead event, not including those lost contact for over

Table 1. Summary of 194 glioma samples and patient follow-up.

Grade[1]	Histology	CBTRUS 1973–2001	Tumor source			Patients		Survival time (years)			Age	
			MDA	UAMS	UCI	Observed	Events[2]	Median	0.95 LCL	0.95 UCL	Median	Max
1	O	10 yrs at 53.2%	18	0	0	18	9	9.67	4.88	NA	34	59
2	AO	4 yrs at 48.0%	17	3	7	27	13	7.12	2.73	NA	45	60
3	AA	2 yrs at 44.2%	42	5	2	49	33	3.08	1.67	NA	38	53
4	GBM	1 yr at 29.1%	66	21	13	100	92	0.81	0.63	1.00	53	83

Notes: [1]Histology treated as a four-level numeric variable; [2]Death events were updated to April 2010.
Abbreviations: O, oligodendroglioma; AO, anaplastic oligodendroglioma; AA, anaplastic astrocytoma; GBM, glioblastoma multiforme; LCL, lower 95% confidence limit; UCL, upper 95% confidence limit.

5 years), almost mature for AA with death event of 67%, and a 49% death event for AO_O group. Genes selected for inclusion in this study were based on published pilot prognosis data with rationale for their involvement in glioma malignance and resistance as detailed in the Introduction section. We included two main clinical variables that correlated with survival: age and recurrence at the time of surgical removal of the tumor that were also included in our earlier prognostic study for GBM and AA.[6] The 1p/19q deletion data for oligodendralgial tumors is not included in modeling due to lack of such information for large number of the studied subjects.

Tumor specimens and tumor cDNA samples

The cDNA samples from MDACC have been used by us and others in published studies, with indication of sample quality control, RNA, and details on cDNA sample processing.[5,6,23] The RNA samples of gliomas from UAMS and UCI were processed from 2–4 mm^3 snap-frozen tumor pieces using Ultraspec (Biotecx Laboratories, Houston) with an initial homogenization by passing the tissue through a 20-Gauge needle attached to a 1 ml syringe, then RNA extraction following manufacturer's protocol. The integrity of RNA samples were examined by a RNA gel for the presence of 18S and 28S RNA bands. The cDNA was reverse transcribed from an aliquot of RNA (0.5–2 μg) in a 10 μl reaction using 100 Units of Supercript II reverse transcriptase (Invitrogen, Carlsbad), 10 pmol of poly(dT) 20VN primer, and other components provided in the kit following the manufacturer's protocol. The cDNA synthesis reaction was diluted 30 times with 10 mM Tris.HCl (pH 7.5), and an aliquot of 4 μl diluted cDNA was quantified for *ACTB*. Based on *ACTB* quantity, we further diluted the cDNA, eg, 1000 copy number per 4 μl per quantification reaction, for efficient use of the tumor cDNA samples.

Gene expression data for prognosis study

We used Ziren® Human Real-Time AqRT-PCR Standard-1001 (Ziren Research LLC, Irvine) to quantify *ABCG2*, *BMI1*, *MELK*, *MSI1*, *PROM1*, and *ACTB,* and AqRT-PCR Standard-1020 to quantify *PAX6*, *PTEN*, *VEGFA*, and *ACTB* for the entire set of cDNA included in this study. *IGFBP2* and *ACTB* were quantified with AqRT-PCR Standard-1009 for non-GBM samples. Quantification of *CDK4*, *EGFR*, and *MMP2* was carried out in UAMS and UCI on two sets of cDNA samples based on the same single standard containing these three genes, but not the reference gene *ACTB*. For these three genes, the data was a relative ratio to *ACTB*. In order to combine the two sets of relative quantitative data in to a single prognosis model, the data in the 2nd set of samples (mainly composed of GBM) was adjusted by the mean of fold difference between the 2nd and the 1st quantification in a set of the same cDNA samples from the 1st set.

We used FAST-START DNA Master SYBR Green I mix (Roche, Indianapolis) in real-time PCR using a LightCycler 2.0 real-time instrument (Roche) or step one real-time PCR instrument (Applied Biosystems, Foster City). The primers are designed to amplify all transcription variants for genes, including *EGFR*, *MMP2*, *PAX6*, and *VEGFA*. Primer sequence information and PCR parameters are available upon request to Ziren Research (www.zirenresearch.com). The quantification for each gene was repeated 2–4 times for each cDNA sample, and the mean value was used for calculating the ratio of marker gene to *ACTB*. In our prior glioma prognosis study,[5] we have shown that *ACTB* (called β-Actin there) is a fair reference gene to normalize the marker gene expression among different glioma samples, and we included some of these prior quantified data into this study dataset.

Statistical analyses

From our earlier two studies, we found that modeling glioma prognosis by dichotomizing patients based on recursive partitioning of raw gene ratios produced a model with a higher and biased R^2 value than treating the gene expression data as continuous variables after log transformation.[5,6] Thus, in this study, we applied multivariate Cox Proportional Hazard (PH) models to estimate the prognoses of GBM, AA, and AO_O using log-scaled ratios (logRatios) of marker vs. *ACTB* expression quantities. The application of log transformation also avoids outliers on the right and ensures a reliable result. To avoid taking log of zero we used an offset selected to be large enough (such as 0.01) to avoid outliers on the left.

We computed the univariate model coefficients for three glioma groups graded based on tumor malignancy (GBM as grade 4, AA as grade 3, AO_O as

grade 1/2) from computer output for the models with grade-gene interactions. The model is represented as follows: log HR = b1*gene + b2*g1 + b3*g2 + b4*g1*gene + b5*g2*gene. b1, b2, b3, b4, and b5 are model coefficients; gene is the log(gene + 0.01) value; g1 is a dummy variable coded 1 for grade 3, 0 otherwise; and g2 is a dummy variable coded 1 for grade 4, 0 otherwise. We also assessed whether the log hazard ratios change with grade by a test for significance of interaction of genes with grade. Kaplan-Meier survival analysis was performed for a highly significant prognosis factor *MSI1* dichotomized at the overall median expression in a total of 194 glioma samples including GBM, AA, and AO_O. We used Spearman Rank Correlation Test to analyze gene expression correlations, and Wilcoxon Rank Sum Test to analyze the difference on gene expressions among the three tumor grades (GBM, AA, and AO_O). All statistical analyses were performed using S-PLUS 2000 computer software (MathSoft Inc, Seattle, WA).

Results

The prognostic effect of age and recurrence in gliomas

As shown in Table 1, survival time for patients with GBM is much shorter and less varied than those with other types of gliomas. We performed Cox PH analyses for GBM, AA, AO_O to assess the prognostic value of age and recurrence. As shown in Model 1 in Tables 2 and 3, recurrence, not age, is a significant predictor of poor survival for AA, and none of the clinical variables are significant in GBM and AO_O.

To substantiate above finding, we combined all glioma grades in a Cox PH analysis with treatment of histology as a four-level numeric variable, as shown in Table 1. The result revealed that GBM histology has a significant power in prediction of poor survival versus non-GBM with Log (HR) = 1.23 ($P < 0.0001$), adjusted for other clinical variables (age and recurrence). There is a significant correlation between grade and age (R = 0.42, $P < 0.00001$), which is consistent with a bias in age with older population in GBM compared to non-GBM (see Table 1). The hazard ratio (HR) for a 20-year increase in age was 1.25 with $P = 0.092$ in all gliomas (1.41, 2.09, and 1.13 for AO_O, AA, and GBM, respectively), while HR for age did not vary significantly with grade ($P = 0.27$).

As shown in Tables 2 and 3, tumor recurrent status (recurrence) at time of sample collection, which is a binary variable versus non-recurrence, is an unfavorable prognostic factor in multivariable models for AA, but not for other glioma grades. Consistent with the finding, the HR for recurrence varied significantly with grade ($P = 0.025$): 1.83, 3.26, and 0.92 for AO_O, AA, and GBM, respectively.

Further analysis was carried out to exam the prognostic effect of grade-age and grade-recurrence interactions on patient's OS. The data showed that there is a significant GBM–recurrence interaction with Log (HR) = −1.2606, $P = 0.0024$, indicating that recurrence has a different function in predicting survival for GBM and non-GBM, which can be explained by the fact that GBM is the highest malignancy and re-operation is beneficial to OS, but recurrence in non-GBM is related to tumor progression into a higher malignancy thus a poor prognostic factor.

Prognostic effect of gene expression variable in univariate models of gliomas

We performed univariate Cox PH assay for each gene separately for GBM, AA, and AO_O. With consideration of sample size and based on their common histological features, we combined AO and O cases in this study, We plotted the hazard ratio (HR) vs. gene expression logRatio curves based on the univariate Cox PH model. As shown in Figure 1 and Table 4, from the 13 genes, only *PTEN* showed the same decreasing curve and consistently an unfavorable value for Log (HR) in all three glioma types. The other 12 genes showed different effects on prognosis in different glioma subsets classified based on histology.

In GBM, the HR curves for five pathway associated genes (*CDK4*, *VEGFA*, *EGFR*, and *MMP2*) are in general not altered by gene expression levels, with the univariant coefficient Log (HR) being around zero. In contrast, all these genes have either favorable (*MMP2*, *EGFR*) or unfavorable (*CDK4*, *VEGFA*) prognostic values in univariate models of AO_O and AA (except *EGFR*). Although over-expressed in GBM, *IGFBP2* expression showed a favorable effect on prognosis for GBM. In contrast, *IGFBP2* expression in non-GBM gliomas has an unfavorable prognostic value. In accordance with these findings, there is significant interaction of *IGFBP2* with glioma grades (Table 4).

Table 2. Cox PH glioma models based on clinical and gene expression variables.[a]

Histology	Model 1			Model 2			Model 3			Model 4			Model 5		
	GBM	AA	AO_O	GBM	AA	AO_O	GBM	AA	AO_O	GBM	AA	AO_O	GBM	AA	AO_O
R square[b]	0.5%	13.6%	5.3%	2.4%	32.7%	57.6%	11.2%	22.0%	35.6%	14.2%	45.2%	62.6%	11.1%	33.6%	58.1%
P value[c]	0.774	**0.0277**	0.299	0.983	0.059	**0.0016**	0.106	0.095	**0.0071**	0.358	**0.0285**	**0.0076**	0.468	0.142	**0.0054**
Cases	100	49	45	100	45	34	100	49	45	100	45	34	100	45	34

Notes: [a]An offset of 0.01 was chosen for log-transformed mRNA ratios to avoid outliers on the left; [b]as R^2, an index of the Cox PH model showing the percentage of variation in survival explained by the model; [c]Likelihood ratio test compares each model to a null model (one with no covariates) to test whether all of the model coefficients are simultaneously equal to zero. The cutpoints used for significance is $P < 0.05$. Model 1: Two clinical variables (patient age, recurrence status); Model 2: Two clinical variables plus previous studied 7 (GBM) or 8 (non-GBM) genes (*CDK4*, *EGFR*, *MMP2*, *VEGF*, *PAX6*, *PTEN*, *RPS9*, and *IGFBP2* for non-GBM models); Model 3: Two clinical variables plus 5 stem cell associated genes (*ABCG2*, *BMI1*, *MELK*, *MSI1*, *PROM1*); Model 4: Two clinical variables plus genes included in Models 2 and 3; Model 5: omitting clinical variables from Model 4.

The HR-gene expression curve and Log (HR) values revealed consistently that *BMI1* expression had an unfavorable prognostic effect for GBM, but favorable for AA and AO_O, and *MELK* expression a favorable for GBM, but unfavorable for AA and AO_O. *PROM1* expression was shown to be unfavorable prognostic factor for AA, but favorable for AO_O and GBM. In contrast, the expressions of *ABCG2*, *PAX6* and *RPS9* are favorable prognostic factors for both AA and GBM, but unfavorable for AO_O.

Based on Kruskal-Wallis Rank Sum test, *MSI1* is one of the two genes (the other being *CDK4*) that is not differentially expressed among the three types of gliomas with $P > 0.05$. However, *MSI1* expression shows a strong opposing effect on prognosis for GBM and AO_O that is out of displaying range of HR-gene expression plot, with a significant interaction to grades shown in Table 4. Further analysis of *MSI1* prognosis function was carried out in univariate models as continuous and dichotomized variables shown below.

Establishment of multivariate models for GBM, AA, and AO_O

The overall analyses revealed different prognosis effects of the same set of genes in different glioma histopathology classifications, stressing the need of establishing multivariate prognosis models based on histology-classified glioma groups. Using a single data set with logRatios of 13 gene expressions to *ACTB*, two clinical variables (numeric data for patient age, and binary data for recurrence), and the patients OS time, we performed Cox PH regression analyses for GBM, AA and AO_O. We analyzed different combinations of variables by generating sub-models with two clinical variables (Model 1), with addition of the 8 pathway related genes from our previous study (Model 2), or the 5 stem cell associated genes (Model 3), and addition of all 13 genes (Model 4). We also examined a model only with the 13 gene expression variables (Model 5) to assess the prognostic significance of the genes independent of patient's age and recurrence of the tumor. Table 2 summarizes the R^2 and *P* value from a likelihood ratio test for the models. Table 3 shows each variable's log hazard ratio and the statistical significance of each model. The individual models were generated to compare the effect of adding the 5 stem cell associated markers to the original model with 8 cancer pathway associated genes, in

Table 3. Estimated parameter values, their estimated standard errors and the *P*-values in a multivariate Cox models shown in Table 2.

Histology		GBM			AA			AO_O		
Variable		**Coef[a]**	**Se (Coef)**	***P* value[d]**	**Coef**	**Se (Coef)**	***P* value**	**Coef**	**Se (Coef)**	***P* value**
Model 1	Age[b]	0.004	0.007	0.57	0.031	0.018	0.08	0.017	0.020	0.39
	Recur[c]	–0.057	0.238	0.81	1.043	0.395	**0.008**	0.629	0.482	0.19
Model 2	Age	0.004	0.008	0.65	0.018	0.023	0.43	0.033	0.032	0.30
	Recur	–0.025	0.249	0.92	0.986	0.520	0.058	–0.716	0.943	0.45
	PAX6	–0.058	0.190	0.76	–0.396	0.419	0.34	0.276	0.651	0.67
	PTEN	–0.072	0.185	0.70	–0.129	0.674	0.85	0.273	0.509	0.59
	VEGFA	–0.003	0.120	0.98	0.319	0.504	0.53	0.518	0.501	0.30
	CDK4	–0.032	0.109	0.77	0.835	0.391	0.033	0.879	0.587	0.13
	EGFR	0.038	0.077	0.62	0.216	0.178	0.23	–1.670	1.227	0.17
	MMP2	0.049	0.221	0.82	–2.339	0.817	**0.004**	–0.643	0.466	0.17
	RPS9	–0.007	0.182	0.97	–0.245	0.321	0.45	0.623	0.405	0.12
	IGFBP2	n/a	n/a	n/a	0.461	0.298	0.12	2.054	0.896	**0.022**
Model 3	Age	0.007	0.008	0.40	0.033	0.020	0.10	–0.003	0.024	0.90
	Recur	–0.262	0.259	0.31	1.158	0.446	**0.009**	0.890	0.582	0.13
	ABCG2	–0.198	0.1529	0.19	–0.381	0.230	0.10	0.188	0.290	0.52
	BMI1	0.161	0.274	0.56	0.376	1.395	0.79	–0.705	1.562	0.65
	MELK	–0.331	0.237	0.16	0.371	0.352	0.29	1.095	0.528	0.038
	MSI1	1.546	0.608	**0.011**	0.329	0.477	0.49	–3.123	1.342	**0.020**
	PROM1	–0.036	0.214	0.87	0.857	0.885	0.33	–0.263	0.723	0.72
Model 4	Age	0.009	0.008	0.31	0.054	0.026	0.039	0.037	0.042	0.38
	Recur	–0.358	0.276	0.19	1.937	0.694	**0.005**	–1.458	1.312	0.27
	ABCG2	–0.185	0.111	0.10	–0.695	0.379	0.066	–0.472	0.562	0.40
	BMI1	0.208	0.304	0.49	0.649	1.801	0.72	2.004	2.671	0.45
	MELK	–0.562	0.280	0.045	–0.716	0.453	0.11	0.753	1.079	0.48
	MSI1	1.826	0.637	**0.004**	1.115	0.659	0.09	–1.540	1.672	0.36
	PROM1	0.074	0.315	0.82	1.112	1.001	0.27	0.365	2.244	0.87
	PAX6	–0.041	0.206	0.84	–0.087	0.502	0.86	0.927	0.865	0.28
	PTEN	–0.059	0.200	0.77	0.325	0.738	0.66	0.197	0.763	0.80
	VEGFA	–0.072	0.131	0.58	0.523	0.570	0.36	0.271	0.874	0.76
	CDK4	–0.024	0.104	0.82	1.435	0.472	**0.002**	0.945	0.780	0.23
	EGFR	0.015	0.075	0.84	0.006	0.220	0.98	–1.704	1.235	0.17
	MMP2	–0.001	0.239	1.00	–2.944	0.964	**0.002**	–0.331	0.581	0.57
	RPS9	0.346	0.223	0.12	–0.410	0.352	0.24	0.518	0.444	0.24
	IGFBP2	n/a	n/a	n/a	0.276	0.298	0.35	1.883	1.144	0.10
Model 5	*ABCG2*	–0.163	0.105	0.12	–0.141	0.304	0.640	–0.340	0.476	0.48
	BMI1	0.185	0.299	0.54	–0.620	1.811	0.730	0.749	1.757	0.67
	MELK	–0.394	0.260	0.13	–0.414	0.433	0.340	0.341	0.941	0.72
	MSI1	1.292	0.575	0.025	0.662	0.637	0.300	–1.854	1.565	0.24
	PROM1	0.078	0.322	0.81	1.219	1.016	0.230	0.158	2.049	0.94
	PAX6	–0.044	0.201	0.83	–0.312	0.468	0.510	0.13	0.569	0.82
	PTEN	–0.102	0.199	0.61	0.152	0.765	0.840	0.139	0.508	0.78
	VEGFA	–0.054	0.128	0.67	0.727	0.530	0.170	0.542	0.663	0.41
	CDK4	0.007	0.101	0.95	1.263	0.516	**0.014**	1.085	0.809	0.18
	EGFR	0.015	0.076	0.84	–0.012	0.212	0.960	–1.106	1.113	0.32
	MMP2	0.032	0.244	0.90	–2.242	0.924	**0.015**	–0.295	0.473	0.53
	RPS9	0.259	0.217	0.23	–0.217	0.345	0.530	0.639	0.428	0.14
	IGFBP2	n/a	n/a	n/a	0.399	0.293	0.170	2.153	1.133	0.06

Notes: [a]Log hazard ratio (HR) for one unit marker value increase; [b]Numeric data between the time of tumor been removed to the time of patient birth; [c]Binary data regarding de novo or recurrence of the tumor; [d]The *P*-value threshold of 0.010 to control the false discovery rate at 5%, and 0.024 at 10%, according to Pounds and Morris 2003.[42]

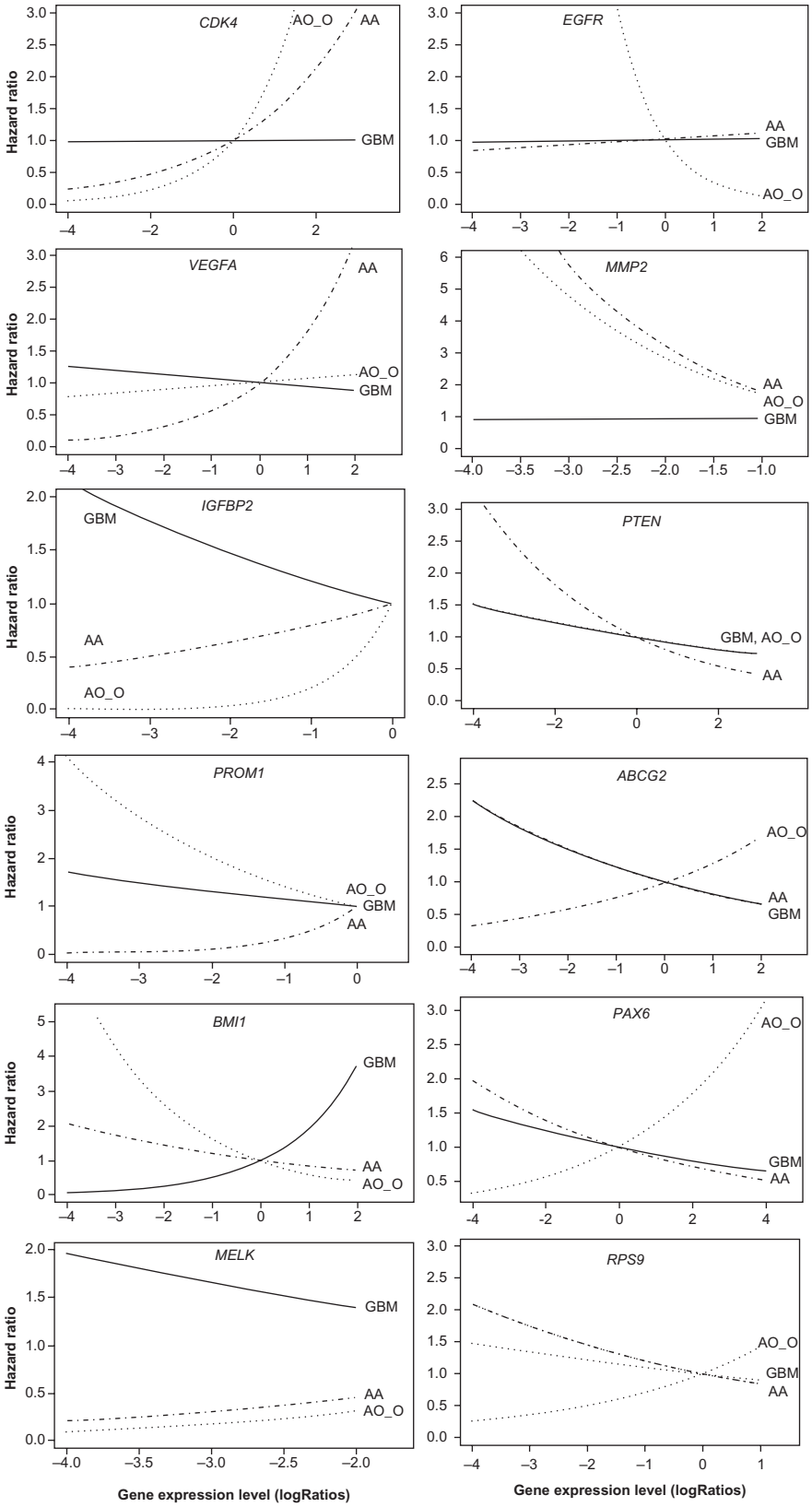

Figure 1. Hazard ratio vs. gene expression logRatios curves for GBM, AA, and AO_O based on the univariate Cox PH model. The hazard ratio for a particular marker corresponding to each of the three grades was computed using a Cox PH model with 3 terms (grade, marker, and grade-marker interaction), for detail statistical analyses see Method. The hazard ratios are shown using zero as the comparator value. A decreasing curve indicates a favorable prognostic effect from the gene expression. In contrast, an increasing curve indicates an unfavorable prognostic effect, while a flat curve signifies no prognostic effect from the gene expression.

Table 4. Log-hazard ratios computed from the univariate model coefficients for each glioma group (AO_O as grade 1/2, AA as grade 3 and GBM as grade 4).

GENE	Log-hazard ratios			*P*-value
	AO_O	AA	GBM	
PROM1	–0.35	1.14	–0.13	0.44
ABCG2	0.27	–0.20	–0.20	0.053
MELK	0.60	0.40	–0.17	0.097
BMI1	–0.49	–0.18	0.66	0.74
MSI1	–2.67	0.21	1.15	**0.0011**
PAX6	0.19	–0.15	–0.11	0.44
PTEN	–0.11	–0.30	–0.10	0.91
VEGFA	0.06	0.59	–0.06	0.19
CDK4	0.75	0.37	0.00	0.12
EGFR	–1.14	0.05	0.01	0.091
MMP2	–0.53	–0.59	0.03	0.24
IGFBP2	1.58	0.23	–0.19	**0.0046**
RPS9	0.35	–0.19	–0.10	0.14

Note: *P* values are from a test for significance of interaction of genes with grade.

order to gain more biological insights on the function of cancer stem cell associated gene expression to residual risk of glioma. We report below a summary of the results from each model for each glioma type.

GBM model

In a multivariate model including 100 GBM patients with 92 events on OS, the two clinical variables (age and recurrence) did not show significant prognostic value and failed to produce a significant prognosis model (R^2 = 0.5%, P = 0.77) (GBM model 1). We have shown in our prior glioma prognosis study[6] that *IGFBP2* expression is significantly correlated with GBM histology and four of the 8 pathway associated genes (*MMP2*, *VEGFA*, *RPS9*, and *PAX6*) and lacks a significant prognostic value in a multivariate model with these variables, thus it is not included in GBM prognosis modeling in this study in an attempt to increase the ratio for events to variables number. We analyzed prognostic model for GBM by including expression variables of *CDK4*, *EGFR, MMP2*, *VEGFA*, *PAX6*, *PTEN*, and *RPS9*, which made no significant improvement with a model R^2 = 2.4% (P = 0.983) (Table 2) and none of the variables showed a significant prognostic value (Table 3). The three genes *PTEN*, *RPS9*, *CDK4* that showed significant prognostic value in a GBM_AA mixed model with 41 GBM and 43 AA cases in our earlier study[6] failed to show prognostic significance in the model with 100 GBM cases.

In the GBM model 3 with 5 stem cell associated gene expressions, a marginally significant improvement was seen in the prognosis model (R^2 = 11.2%, P = 0.106). The expression of the genes *PROM1* (= 0.41) and *ABCG2* (R = –0.31) showed a significant correlation ($P < 0.00001$) with grade, R = 0.41, –0.31, respectively. Both genes showed significant prognostic value in univariate models based on all 194 gliomas —*PROM1*, Log (HR) = 0.188 (P = 0.015), and *ABCG2*, Log (HR) = –0.274 ($P < 0.0001$). However, when adjusted with dichotomous GBM/non-GBM variables, both genes lost their prognostic value for GBM, suggesting that *ABCG2* and *PROM1* prognosis values are confounded by glioma grade. In contrast, with a lack of correlation to glioma histology, *MSI1* expression was found to be a statistically significant negative prognostic factor for GBM. The unfavorable prognostic value of *MSI1* to GBM was seen with or without the inclusion of clinical variables and/or the 7 pathway related genes.

AA model

Recurrence was found to be an unfavorable prognostic factor for the AA group with 2/3 mature patient's survival information, and together with age variable showed prognostic significance in the AA model 1 (R^2 = 13.6%, P = 0.0277). The addition of the 8 cancer pathway related genes (*CDK4*, *EGFR, MMP2*, *VEGFA*, *PAX6*, *PTEN*, *RPS9*, and *IGFBP2*) or the 5 stem cell associated genes, improved the R^2 of the AA model (AA Model 2, R^2 = 32.7% and AA model 3, R^2 = 22.0%) with marginal significance on the likelihood ratio test. The AA Model 4 with two clinical and 13 gene expression variables together achieved likelihood significance with an explanation of 45.2% of the variation in OS of the patients. In the AA Model 5 excluding the two clinical variables and with the remaining 13 gene expression variables the R^2 dropped to 33.6% with a lack of significance in the likelihood ratio test (P = 0.142), consistent with the fact that recurrence is a strong unfavorable prognostic factor for AA.

In AA, the expression of *CDK4* and *MMP2* showed independent significant prognostic values; *CDK4*, as an unfavorable prognostic factor and *MMP2* as a favorable prognostic factor, based on the Log (HR) and *P* values shown in Table 3. None of the stem cell associated genes showed prognostic significance in the

multivariate models of AA. Although the R^2 increased from 13.6% in the model with two clinical variables to 22.0% by adding the 5 stem cell associated gene variables, the model was not significant in the likelihood ratio test. There are multiple pairwise correlations among these genes, which explain the lack of individual significance but an overall improvement of the model in explaining the variation in survival.

AO_O Model

In the multivariate model of AO_O with 48%–50% events in each grade, neither of the two clinical variables showed prognostic significance and failed to produce a significant prognosis model (R^2 = 5.3%, $P = 0.299$). However the addition of 8 pathway related genes (AO_O Model 2) improved the model R^2 value from 5.3% to 57.6%, and the model was significant based on a likelihood ratio test ($P = 0.0016$). Including the 5 stem cell associated gene expression variables also greatly improved the model R^2 values from 5.3% to 35.6%, with a significant P value 0.007 in likelihood ratio test of the model. There was a further increase of model R^2 to 62.6% (AO_O Model 4) with the addition of all 13 gene expressions. The exclusion of clinical variables only decreased the R^2 by 4.5%, consistent with a R^2 of 5.3% in a model with only the clinical variables (AO_O Model 1). Thus the 13 gene expression variables we included in this study have significant prognostic value for oligodendroglial tumors.

In the AO_O prognosis model, *IGFBP2* expression showed an independent significant unfavorable prognosis effect for oligodendraglial tumors, with Log (HR) = 2.1 ($P = 0.022$) in AO_O model 2, in which the number of cases dropped from 45 to 34 due to lack of cDNA samples for 9 cases. In a total of 45 AO_O patients, *MSI1* expression showed a favorable prognostic value with Log (HR) = –3.1 ($P = 0.020$), while the expression of *MELK* an unfavorable prognostic value, with Log (HR) = 1.1 ($P = 0.036$), independent of clinical variables and the other 4 stem cell associated gene expression variables.

Comparison of models with different sets of variables

We compared Model 4 to Model 1 to assess the contribution of the whole set of genes to the model (with just the two clinical variables. The log-ratio P-values for comparing these models are: <0.0001 for AO, 0.0006 for AA, and 0.25 for GBM, indicating that in aggregate the whole 13 genes contribute significantly for AA and AO_O models, but not for GBM. We compared Model 4 to Model 2 to assess the contribution of the 5 stem cell markers over the cancer pathway associated genes. The log-ratio P-values for comparing these models are: 0.64 for AO, 0.093 for AA, and 0.025 for GBM. Thus, in aggregate the 5 stem cell markers contribute significantly only for GBM. The contribution of the 5 stem cell markers improved predictive accuracy for AA and AO_O, but not to a level of statistical significance. So for AA and AO_O, the models with the new (5 stem cell markers) + previously included (the 8 cancer pathway associated markers) genes were significantly better than models with no genes but not significantly better than models with just the previously included genes.

Differential prognostic impact of *MSI1* expression in GBM, AA, and AO_O

The results from this study revealed an interesting finding about the prognostic impact of *MSI1* expression in glioma; an unfavorable prognostic value for GBM, no effect for AA, and a favorable prognostic value for AO_O patients. Consistent with the results from the multivariate Cox PH analysis, *MSI1* expression as a continuous variable has a significant unfavorable prognostic value in the univariate model for GBM [Log (HR) = 1.23, R^2 = 5.2%, P = 0.02], significant favorable prognostic value for AO_O [Log (HR) = –2.83, R^2 = 19.2%, P = 0.002], and a lack of prognosis value for AA [Log (HR) = 0.21, $R^2 = 0.4\%$, $P = 0.65$]. We further explored if patient's survival can be distinguished by dichotomizing based on a biologically relevant threshold of *MSI1* expression in gliomas. We set the threshold at an overall median of 0.0012 in the raw (absolute ratio of *MSI1* to *ACTB*) data from all 194 gliomas, and analyzed the hazard ratio (HR) and the survival variation by Kaplan-Meier survival curves. As shown in Figure 2, in contrast to the results from treating *MSI1* expression as a log-scaled continuous variable, dichotomizing raw *MSI1* expression ratios failed to separate GBM patients with a significant difference in survival. In agreement with the data in log-scale as a continuous variable, higher *MSI1* expression showed an unfavorable prognostic value with HR = 1.3, 95% CI = (0.9, 2.0).

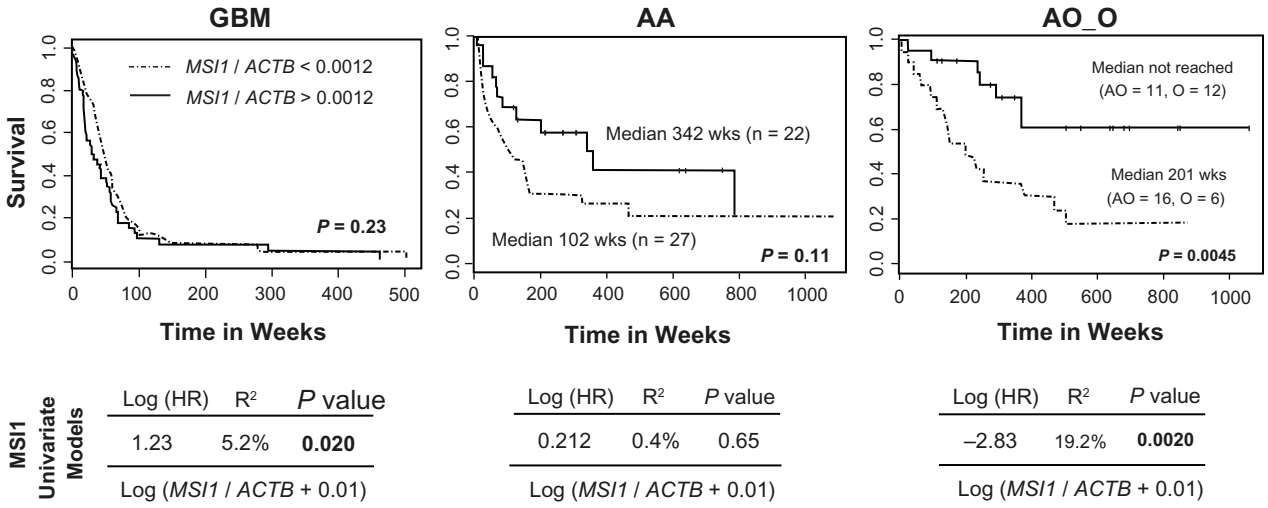

Figure 2. Opposing effect of *MSI1* in prognosis for GBM and oligodendroglia tumors. **Upper panel** shows Kaplan-Meier survival curves for GBM, AA, and AO_O based on absolute ratio of *MSI1* to *ACTB* dichotomized at the overall median of 0.0012 for all 194 gliomas. **Bottom panel** shows log scaled *MSI1* univariate models for GBM, AA and AO_O.

Dichotomizing raw *MSI1* expression ratios was able to separate AO_O patients with a significant difference in survival time ($P = 0.0045$) and showed a favorable prognostic value with HR = 0.3, 95% CI = (0.1, 0.7). The effect appears to be independent of AO and O histology, as the distribution of these two groups is not skewed with *MSI1* expression. In agreement with non-significant positive prognosis value of *MSI1* expression as a continuous variable, dichotomized *MSI1* variable in AA also showed a non-significant positive effect on prognosis ($P = 0.11$) with HR = 0.6, 95% CI = (0.3, 1.2).

Different molecular signatures in GBM, AA, and AO_O based on gene expression correlation analyses

Using a Spearman rank correlation test, we analyzed the pair-wise correlation on gene expressions in different glioma types graded based on histology and survival rate: 4 for GBM, 3 for AA, 2 for AO, and 1 for O. As shown in Table 5, among the 5 stem cell associated genes, there is a lack of significant correlation in GBM. A significant positive correlation between *BMI1–ABCG2* (R = 0.35, $P = 0.013$) and an unfavorable correlation between *MSI1–MELK* (R = –0.36, $P = 0.011$) was seen in AA. The *BMI1–ABCG2* positive correlation is greater in AO_O (R = 0.50, $P = 0.0005$), but the other genes lack a significant correlation in AO_O.

We further analyzed the correlation between the stem cell associated genes and the 8 previously studied cancer pathway associated genes. As shown in Table 5, three stem cell associated genes (*ABCG2*, *BMI1*, and *MELK*) are positively correlated with *PAX6* or *PTEN* in GBM, AA and AO_O, which have been shown to have a decreased expression in GBM compared to AA or surrounding normal tissues[5] and play tumor suppression functions in GBM-derived cell lines.[24–27] There is a significant positive correlation between the pro-angiogenic gene *VEGFA* and the expression of different stem cell associated genes in different tumor grades; *PROM1–VEGFA* in GBM (R = 0.46, $P < 0.0001$), *MSI1–VEGFA* in AA (R = 0.36, $P = 0.0117$); *MELK–VEGFA* in AO_O (R = 0.39, $P = 0.0074$). There is also a significant positive correlation between *ABCG2–PAX6* and *ABCG2–PTEN* across all glioma histologies. *VEGFA* expression is also significantly positively correlated with *PAX6* (R = 0.48, $P = 0.0008$) and *PTEN* (R = 0.42, $P = 0.004$) in AO_O, but not seen in AA and at a reduced level in GBM.

In GBM, there is a general lack of significant correlation between stem cell associated genes and those directly in control of signaling pathways related to tumor aggressiveness, such as *CDK4*, *EGFR*, *MMP2*, and *IGFBP2*, all were shown to have an increased expression in GBM compared to AA or surrounding normal tissues.[5,6] In contrast, there is a strong significant positive correlation between pro-proliferation gene *CDK4* and several stem cell associated genes (*ABCG2*, *BMI1*, and *MELK*) in AA and AO_O. *MSI1*

Table 5. Spearman rank correlation coefficient matrix with *R* (upper part) and *P* (lower part) values for GBM, AA, and AO_O. (correlations with $P < 0.02$ are in bold face, $P \leq 0.001$ are in box).

GBM	*PROM1*	*ABCG2*	*MELK*	*BMI1*	*MSI1*	*PAX6*	*PTEN*	*VEGFA*	*CDK4*	*EGFR*	*MMP2*	*IGFBP2*	*RPS9*
PROM1	1.00	0.07	0.20	0.19	0.09	**0.29**	**0.27**	**0.46**	–0.01	0.12	0.12	–0.07	0.00
ABCG2	0.4794	1.00	0.17	0.08	–0.13	**0.41**	**0.34**	–0.04	0.03	0.18	0.11	–0.22	**0.32**
MELK	0.0466	0.0986	1.00	–0.02	0.23	**0.28**	0.13	–0.03	0.11	0.12	–0.14	–0.12	**0.28**
BMI1	0.0587	0.4090	0.8160	1.00	0.15	0.15	0.13	–0.03	0.14	0.14	**0.27**	–0.02	–0.07
MSI1	0.3686	0.1894	0.0207	0.1480	1.00	–0.06	–0.02	0.17	0.06	0.02	–0.06	–0.10	–0.13
PAX6	**0.0030**	**0.0000**	**0.0041**	0.1494	0.5584	1.00	**0.52**	**0.32**	–0.05	0.16	–0.12	**–0.35**	**0.34**
PTEN	**0.0061**	**0.0006**	0.2041	0.1822	0.8138	**0.0000**	1.00	0.22	–0.03	0.11	0.02	**–0.43**	**0.27**
VEGFA	**0.0000**	0.6945	0.7305	0.7918	0.0927	**0.0011**	0.0274	1.00	–0.04	0.13	0.01	–0.07	0.03
CDK4	0.9298	0.7520	0.2555	0.1641	0.5444	0.6444	0.7615	0.6906	1.00	0.14	0.07	0.18	0.18
EGFR	0.2366	0.0769	0.2526	0.1732	0.8798	0.1094	0.2868	0.1890	0.1652	1.00	0.31	0.13	**0.34**
MMP2	0.2343	0.2873	0.1649	**0.0076**	0.5415	0.2209	0.8212	0.9561	0.4805	0.0020	1.00	**0.33**	0.07
IGFBP2	0.4883	0.0255	0.2381	0.8110	0.3286	**0.0004**	**0.0000**	0.4862	0.0680	0.2065	**0.0009**	1.00	–0.12
RPS9	0.9995	**0.0010**	**0.0044**	0.4818	0.1878	**0.0006**	**0.0072**	0.7762	0.0805	**0.0006**	0.4866	0.2272	1.00
AA	***PROM1***	***ABCG2***	***MELK***	***BMI1***	***MSI1***	***PAX6***	***PTEN***	***VEGFA***	***CDK4***	***EGFR***	***MMP2***	***IGFBP2***	***RPS9***
PROM1	1.00	0.08	–0.12	0.28	0.08	–0.19	–0.18	0.16	0.23	0.04	0.13	0.30	–0.14
ABCG2	0.6024	1.00	0.19	**0.35**	–0.01	**0.49**	**0.49**	–0.07	**0.47**	0.07	0.05	–0.18	0.11
MELK	0.4299	0.1896	1.00	0.33	**–0.36**	–0.03	**0.37**	0.17	**0.47**	0.23	0.21	0.29	–0.17
BMI1	0.0475	**0.0129**	0.0210	1.00	0.05	0.25	**0.41**	0.28	**0.49**	0.06	0.27	0.19	–0.02
MSI1	0.6051	0.9555	**0.0106**	0.7505	1.00	0.14	0.13	**0.36**	0.08	**0.59**	0.07	–0.27	0.31
PAX6	0.1886	**0.0003**	0.8638	0.0785	0.3253	1.00	**0.51**	0.13	0.14	0.15	–0.14	–0.20	0.04
PTEN	0.2229	**0.0003**	**0.0085**	**0.0038**	0.3652	**0.0002**	1.00	0.18	0.19	0.08	–0.07	–0.12	0.09
VEGFA	0.2661	0.6443	0.2388	0.0540	**0.0117**	0.3818	0.2244	1.00	0.00	**0.41**	0.03	0.07	0.02
CDK4	0.1311	**0.0010**	**0.0013**	**0.0007**	0.6136	0.3707	0.2082	0.9784	1.00	–0.22	0.33	0.15	–0.01
EGFR	0.8033	0.6276	0.1322	0.6841	**0.0000**	0.3231	0.5821	**0.0051**	0.1534	1.00	0.12	–0.15	0.16
MMP2	0.4011	0.7553	0.1737	0.0685	0.6299	0.3560	0.6593	0.8318	0.0256	0.4384	1.00	0.32	0.02
IGFBP2	0.0420	0.2440	0.0510	0.2070	0.0747	0.1812	0.4184	0.6591	0.3212	0.3169	0.0349	1.00	**–0.40**
RPS9	0.3556	0.4837	0.2568	0.8865	0.0399	0.7833	0.5367	0.9029	0.9245	0.2975	0.8764	**0.0072**	1.00
AO_O	***PROM1***	***ABCG2***	***MELK***	***BMI1***	***MSI1***	***PAX6***	***PTEN***	***VEGFA***	***CDK4***	***EGFR***	***MMP2***	***IGFBP2***	***RPS9***
PROM1	1.00	0.22	0.25	0.24	0.15	0.36	**0.40**	0.23	0.15	**0.36**	–0.08	0.12	0.32
ABCG2	0.1555	1.00	0.22	**0.50**	–0.21	**0.42**	**0.43**	0.23	**0.50**	0.16	–0.28	–0.04	0.37
MELK	0.1011	0.1498	1.00	0.23	0.08	0.28	0.25	0.39	0.32	0.05	–0.19	–0.09	0.11
BMI1	0.1073	**0.0005**	0.1217	1.00	0.13	0.31	**0.54**	0.30	0.24	**0.35**	–0.03	–0.17	0.12
MSI1	0.3208	0.1725	0.5828	0.4032	1.00	–0.17	–0.09	–0.21	0.06	**0.36**	**0.43**	–0.03	–0.13
PAX6	0.0146	**0.0038**	0.0659	0.0376	0.2537	1.00	0.30	**0.48**	0.31	0.12	–0.33	0.00	**0.60**
PTEN	**0.0069**	**0.0031**	0.0962	**0.0001**	0.5360	0.0448	1.00	**0.42**	0.11	**0.50**	–0.01	–0.02	0.20
VEGFA	0.1233	0.1334	**0.0074**	0.0482	0.1743	**0.0008**	**0.0040**	1.00	0.12	0.28	–0.13	0.01	0.04
CDK4	0.3386	**0.0005**	0.0344	0.1133	0.7162	0.0404	0.4861	0.4208	1.00	0.17	–0.22	–0.06	0.23
EGFR	**0.0150**	0.3043	0.7246	**0.0195**	**0.0141**	0.4471	**0.0004**	0.0644	0.2735	1.00	0.24	0.03	0.07
MMP2	0.6155	0.0600	0.2108	0.8410	**0.0034**	0.0276	0.9418	0.4006	0.1404	0.1057	1.00	–0.08	0.00
IGFBP2	0.4439	0.8176	0.5400	0.2734	0.8558	0.9847	0.8717	0.9620	0.7027	0.8380	0.5846	1.00	0.04
RPS9	0.0641	0.0303	0.5505	0.4951	0.4649	**0.0002**	0.2559	0.8208	0.1938	0.6992	0.9932	0.8051	1.00

expression has a significant positive correlation with *EGFR* (R = 0.59, $P < 0.0001$) and *VEGFA* (R = 0.36, $P = 0.012$) in AA. In AO_O, *MSI1* is also positively correlated with *EGFR* (R = 0.36, $P = 0.0141$), and *MMP2* (R = 0.43, $P = 0.0034$), but not with *VEGFA*.

In agreement with results of an independent study on gene expressions in AA and GBM,[6] there are significant positive correlations for *PAX6–PTEN* in either combined (R = 0.53, $P < 0.0001$) or separate glioma grades with R = 0.52 ($P < 0.0001$) in GBM and R = 0.51 ($P < 0.0001$) in AA. The negative correlations between *PAX6–IGFBP2* (R = –0.35, $P = 0.0004$) and *PTEN–IGFBP2* (R = –0.43, $P < 0.0001$) occur only in GBM, while a positive correlation between *MMP2–IGFBP2* occurs in both GBM and AA, consistent with IGFBP2 up-regulation of *MMP2* expression[28] and PAX6 and PTEN suppression of malignant behaviors of glioma cells.[24,29] Apart from the above indicated gene expression correlations in AA and GBM, positive correlations between *PAX6–VEGFA* ($R = 0.48$, $P = 0.0008$) and *PAX6–RPS9* (R = 0.60, $P = 0.0002$) are seen in AO_O, and specific correlations between *PTEN–VEGFA* (R = 0.42, $P = 0.004$) and *PTEN–EGFR* (R = 0.50, $P = 0.0004$) occur only in AO_O.

Overall data from gene correlation analyses revealed different molecular signatures in GBM, AA, and AO_O, which support above analysis of the prognostic effects of these gene expressions separately in each of these tumor types.

Discussion

This is a study of prognosis based on multiple gene expression information in the tumor specimens for histology classified glioma patients without discriminating their difference on treatments received, because the outlook for patients with malignant gliomas has improved very little since the first randomized prospective clinical trials for malignant astrocytomas published in 1978.[30,31] Only a modest 2.5 month median survival increase has been achieved by adding concomitant temozolamide to radiotherapy after surgery according to a report published in 2005.[32] Our endeavor in modeling glioma prognosis follows the principle of generating reliable models by satisfying the statistical criterion for the ratio between the number of events and variables. Standardized gene quantification ensures data comparability,[7] thereby allowing combination of different cohort data to from a large training set for prognosis study. We selected candidate genes for improvement of R^2 of the prognosis model, based on their individual prognostic value in univariate models on pilot sets of gliomas from published studies, as well as their functional roles in tumor suppression and progression. Our evidence-based selection of genes for modeling prognosis has produced statistically significant prognosis models for AA and oligodendroglial tumors.

Gene expression information in relation to OS of glioblastoma multiforme

In this prognosis study, in order to achieve a large sample size, we combined GBM samples from patients operated between 1990–1997 from one institute (MDACC) and between 2003–2007 from three institutes (MDACC, UAMS, and UCI), regardless of whether patients were subjected to intensive or less intensive treatments. GBM patients, especially the recurrent ones, have been subjected to different concurrent chemo/radiation clinical trials over the last 40 years without much improvement.[30–32] Based on result of this and our prior studies on glioma prognosis, the overall multivariate prognosis model for GBM with 13 gene expressions and 2 clinical variables lacks the power to explain a significant portion of variation on OS. This is unlikely due to sample size issue, given the analysis was carried out based on 100 tumors and 92 events.

Although most of the molecular or clinical variables have shown prognostic values in univariate Cox PH models for a mixed glioma set in our prior studies, their prognostic values were lost when adjusted for GBM in this study. It indicates that GBM diagnosis is in itself a strong prognosis factor, thus genes functionally associated with GBM diagnosis hallmarks (high proliferation index, anaplastic, micro-vascular amplification and/or necrosis) lack independent prognosis values for GBM. Although in general we failed to identify molecular markers for prognosis of GBM, we have identified the *MSI1* gene expression as an unfavorable prognostic factor for GBM. Its effect in prognosis was shown by treating it as continuous variable, as we failed to find a cut point for *MSI1*

expression to dichotomize GBM with a significant difference on OS.

Gene expression information in relation to OS of anaplastic astrocytomas

The study of prognosis for AA has been challenged by a lower incidence (7.5% tumors of neuroepithelial tissue) and a longer survival (2 yrs at 44.0%) compared to GBM. AA is often combined with GBM to increase sample size similar to two of our previous prognosis studies.[5,6] This study provides for the first time a multivariate prognosis model for AA, with main effects from 13 gene expressions (in log scale) and 2 clinical variables, to explain 45.2% of the survival variation with statistical significance. In contrast to GBM and AO_O, recurrence at operation was found to be a significant unfavorable prognostic factor for AA, independent of the 13 gene expression variables included in the multivariable model.

By modeling prognosis in the multivariable model, we identified *CDK4* expression to have an independent significant unfavorable prognosis value for AA, which is consistent with its function in promoting cancer cell proliferation. The other gene with significant prognostic value for AA is *MMP2*, the matrix metallopeptidase gene overexpressed in glioma and functions in promoting glioma cell invasion.[33–35] The data from this study show for the first time that *MMP2* expression is an independent significant favorable prognostic factor for AA. Consistent with the idea that angiogenesis drives tumor progression, *VEGFA* expression has an unfavorable prognostic effect in an univariable model for AA, but not in multivariate models, suggesting *VEGFA* prognostic function is confounded by other prognostic factors in the model.

Gene expression information in relation to OS of oligodendraglial tumors

The same issues that challenge AA prognosis apply to modeling the prognosis of oligodendroglial tumors, which comprise about 8.8% of the overall tumors of neuroepithelial tissue with patient outcomes better than astrocytic gliomas of the same WHO grade. We included mainly those patients operated during 1987–1997 to ensure 50% cases with mature survival information. Based on current follow-up information with the 13 genes and 2 clinical variables, we generated a prognosis model that explains 62.6% of survival variation with statistical significance based on a likelihood ratio test ($P = 0.0076$). The main contributions come from the 8 cancer pathway related genes that markedly improve the model based on 2 clinical variables ($R^2 = 5.3\%$, $P = 0.299$) to a model with $R^2 = 57.6\%$ ($P = 0.0016$). *IGFBP2* expression has an independent unfavorable prognostic value for AO_O. There are multiple pairwise correlations among the remaining 7 genes in AO_O, which probably explains the lack of individual significance but an overall improvement of the model. Based on results from univariate analysis, *MMP2* expression has a favorable effect for prognosis of AO_O, as seen in AA. Different from that in AA or GBM and in contrast to its usual oncogenic role, *EGFR* has a favorable prognostic effect in AO_O. Although the *P* values show a lack of significance, the negative log (HR) values for *MMP2* and *EGFR* are consistent with their favorable prognostic effect in AO_O.

Adding the 5 stem cell associated gene expression variables also greatly improved the model R^2 of 5.3% with the 2 clinical variable to a R^2 of 35.6% ($P = 0.0071$). This improvement apparently comes from independent prognostic values of *MSI1* (favorable) and *MELK* (unfavorable). The other three stem cell associated genes have prognostic values in univariate models of AO_O, but are not significant in the multivariate model, which may be explained by their expression correlations, such as the one between *ABCG2–BMI1* (R = 50%, $P < 0.001$).

Both *MELK* and *IGFBP2* have been shown to promote cell proliferation[14] and invasion[28] in glioma and thus their unfavorable prognostic values are related to differential activation of these two pathways in AO_O. Our finding of a favorable prognosis effect of *MSI1* expression in oligodendroglial tumors is in contrast to its unfavorable prognosis effect in GBM. In contrast to GBM in which we were unable to dichotomize patients based on *MSI1* expression, we were able to set up a threshold, a median level of *MSI1* expression in three glioma sets in combination, to dichotomize patients with AO_O to show a significant difference on OS, regardless of tumor grades (AO or O). This data supports our combining of AO and O into a single set in a prognosis study.

Association of prognostic effects of stem cell associated genes with glioma histology

Results from univariate Cox PH analysis in this study revealed interesting opposing effects from expression of the stem cell associated genes to the prognosis of glioma with different histopathology characteristics. The expressions of *ABCG2*, *MELK*, and the neural stem/progenitor cell-associated *PAX6* showed unfavorable prognostic effects for AO_O, but favorable prognostic effect for AA and GBM. The expressions of other three stem cell associated genes (*MSI1*, *BMI1*, and *PROM1*) showed a favorable prognostic effect for AO_O, but *MSI1* and *BMI1* are unfavorable factors for GBM. In contrast to results on CD133 immunostaining and microarray expression data showing that *PROM1* is an unfavorable prognostic factor for patients with GBM and oligodendroglial tumors[36–41] our data from real-time qRT-PCR quantification in this study set revealed *PROM1* as a favorable prognostic factor for AO_O, lack of prognostic value for GBM, and unfavorable for AA. This discrepancy needs to be further investigated for difference in relation to the detection methods as well as source samples.

Concerns of sample size for AA and AO_O prognostic models

This study generated statistically significant prognosis models that are able to explain the variations on OS for 45% and 63% of patients with AA and AO_O. However, based on the statistical criterion for prognosis modeling, approximately 10 patient outcomes per variable,[2] the model 4 for AA and AO_O with 49 and 34 patients, respectively, needs to be reassessed in a model with proportionate sample size. Although the *P* values from the likelihood ratio test showed significance for both models, there is a need for validating the AA and AO_O prognosis models using an independent test set with increase of sample size.

Acknowledgements

This research was supported in part by Seed Grant from Chao Family Comprehensive Cancer Center at the University of California, Irvine, the UC Irvine Academic Senate Council on Research, Computing and Libraries, and the Arkansas Cancer Research Center at the University of Arkansas for Medical Sciences. We would also like to acknowledge UAMS Tissue Bank for providing tumor specimens, the M.D. Anderson Tumor Registry and Dr. Laura F Hutchins for providing follow-up information for patients from MDACC and UAMS, respectively.

Abbreviations

AqRT-PCR, absolute quantitative reverse transcription-polymerase chain reaction; GBM, glioblastoma multiforme; AA, anaplastic astrocytoma; AO, anaplastic oligodendroglioma; O, oligodendroglioma; PH, proportional hazards; OS, overall survival; HR, hazard ratio.

Disclosure

This manuscript has been read and approved by all authors. This paper is unique and is not under consideration by any other publication and has not been published elsewhere. The authors and peer reviewers of this paper report no conflicts of interest. The authors confirm that they have permission to reproduce any copyrighted material.

References

1. CBTRUS Statistical Report: Primary Brain and Central Nervous System Tumors Diagnosed in the United States in 2004–2006. *Central Brain Tumor Registry of the United States*. 2010.
2. Katz MH. Multivariable Analysis. (Cambridge University Press, Cambridge; 1999).
3. Phillips HS, 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.
4. Verhaak RG, et al. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell*. 2010;17:98–110.
5. Zhou YH, Tan F, Hess KR, Yung WK. The expression of PAX6, PTEN, vascular endothelial growth factor, and epidermal growth factor receptor in gliomas: relationship to tumor grade and survival. *Clin Cancer Res*. 2003;9:3369–75.
6. Zhou YH, Hess KR, Liu L, Linskey ME, Yung WK. Modeling prognosis for patients with malignant astrocytic gliomas: quantifying the expression of multiple genetic markers and clinical variables. *Neuro Oncol*. 2005;7:485–94.
7. Zhou Y-H, Raj VR, Siegel E, Yu L. Standardization of gene expression quantification by absolute real-time qRT-PCR system using a single standard for marker and reference genes. *Biomarker Insights*. 2010;5:79–85.
8. Ignatova TN, et al. Human cortical glial tumors contain neural stem-like cells expressing astroglial and neuronal markers in vitro. *Glia*. 2002;39:193–206.
9. Galli R, et al. Isolation and characterization of tumorigenic, stem-like neural precursors from human glioblastoma. *Cancer Res*. 2004;64:7011–21.
10. Hemmati HD, et al. Cancerous stem cells can arise from pediatric brain tumors. *Proc Natl Acad Sci U S A*. 2003;100:15178–83.
11. Bao S, et al. Glioma stem cells promote radioresistance by preferential activation of the DNA damage response. *Nature*. 2006;444:756–60.
12. Singh SK, et al. Identification of human brain tumour initiating cells. *Nature*. 2004;432:396–401.
13. Hayry V, et al. Stem cell protein BMI-1 is an independent marker for poor prognosis in oligodendroglial tumours. *Neuropathol Appl Neurobiol*. 2008.

14. Nakano I, et al. Maternal embryonic leucine zipper kinase is a key regulator of the proliferation of malignant brain tumors, including brain tumor stem cells. *J Neurosci Res*. 2008;86:48–60.
15. Cui H, et al. Bmi-1 regulates the differentiation and clonogenic self-renewal of I-type neuroblastoma cells in a concentration-dependent manner. *J Biol Chem*. 2006;281:34696–704.
16. Bruggeman SW, et al. Bmi1 controls tumor development in an Ink4a/Arf-independent manner in a mouse model for glioma. *Cancer Cell*. 2007;12:328–41.
17. Dirks P. Bmi1 and cell of origin determinants of brain tumor phenotype. *Cancer Cell*. 2007;12:295–7.
18. Li J, et al. Oncoprotein Bmi-1 renders apoptotic resistance to glioma cells through activation of the IKK-nuclear factor-kappaB Pathway. *Am J Pathol*. 2010;176:699–709.
19. Nakano I, et al. Maternal embryonic leucine zipper kinase (MELK) regulates multipotent neural progenitor proliferation. *J Cell Biol*. 2005;170:413–27.
20. Zhou S, et al. The ABC transporter Bcrp1/ABCG2 is expressed in a wide variety of stem cells and is a molecular determinant of the side-population phenotype. *Nat Med*. 2001;7:1028–34.
21. Sakakibara S, et al. Mouse-Musashi-1, a neural RNA-binding protein highly enriched in the mammalian CNS stem cell. *Dev Biol*. 1996;176:230–42.
22. Kaneko Y, et al. Musashi1: an evolutionally conserved marker for CNS progenitor cells including neural stem cells. *Dev Neurosci*. 2000;22:139–53.
23. Sano T, et al. Differential expression of MMAC/PTEN in glioblastoma multiforme: relationship to localization and prognosis. *Cancer Res*. 1999;59:1820–4.
24. Zhou YH, et al. PAX6 suppresses growth of human glioblastoma cells. *J Neurooncol*. 2005;71:223–9.
25. Su JD, Mayo LD, Donner DB, Durden DL. PTEN and phosphatidylinositol 3'-kinase inhibitors up-regulate p53 and block tumor-induced angiogenesis: evidence for an effect on the tumor and endothelial compartment. *Cancer Res*. 2003;63:3585–92.
26. Li DM, Sun H. PTEN/MMAC1/TEP1 suppresses the tumorigenicity and induces G1 cell cycle arrest in human glioblastoma cells. *Proc Natl Acad Sci U S A*. 1998;95:15406–11.
27. Knobbe CB, Merlo A, Reifenberger G. Pten signaling in gliomas. *Neuro-Oncology*. 2002;4:196–211.
28. Wang H, et al. Insulin-like growth factor binding protein 2 enhances glioblastoma invasion by activating invasion-enhancing genes. *Cancer Res*. 2003;63:4315–21.
29. Davies MA, et al. Adenoviral transgene expression of MMAC/PTEN in human glioma cells inhibits Akt activation and induces anoikis. *Cancer Res*. 1998;58:5285–90.
30. Walker MD, et al. Evaluation of BCNU and/or radiotherapy in the treatment of anaplastic gliomas. A cooperative clinical trial. *J Neurosurg*. 1978;49:333–43.
31. Walker MD, et al. Randomized comparisons of radiotherapy and nitrosoureas for the treatment of malignant glioma after surgery. *N Engl J Med*. 1980;303:1323–9.
32. Stupp R, et al. Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *N Engl J Med*. 2005;352:987–96.
33. Sato H, et al. A matrix metalloproteinase expressed on the surface of invasive tumour cells. *Nature*. 1994;370:61–5.
34. Noel A, et al. Emerging roles for proteinases in cancer. *Invasion Metastasis*. 1997;17:221–39.
35. McCawley LJ, Matrisian LM. Tumor progression: defining the soil round the tumor seed. *Curr Biol*. 2001;11:R25–7.
36. Zeppernick F, et al. Stem cell marker CD133 affects clinical outcome in glioma patients. *Clin Cancer Res*. 2008;14:123–9.
37. Pallini R, et al. Cancer stem cell analysis and clinical outcome in patients with glioblastoma multiforme. *Clin Cancer Res*. 2008;14:8205–12.
38. Beier D, et al. CD133 expression and cancer stem cells predict prognosis in high-grade oligodendroglial tumors. *Brain Pathol*. 2008;18:370–7.
39. Zhang M, et al. Nestin and CD133: valuable stem cell-specific markers for determining clinical outcome of glioma patients. *J Exp Clin Cancer Res*. 2008;27:85.
40. Colman H, et al. A multigene predictor of outcome in glioblastoma. *Neuro Oncol*. 12:49–57.
41. Murat A, 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*. 2008;26:3015–24.
42. Pounds S, Morris SW. Estimating the occurrence of false positives and false negatives in microarray studies by approximating and partitioning the empirical distribution of *P*-values. *Bioinformatics*. 2003;19:1236–42.