A prognostic gene expression signature in infratentorial ependymoma.

(Article begins on next page)

19 April 2024

UNIVERSITÀ DEGLI STUDI DI TORINO

A prognostic gene expression signature in infratentorial ependymoma

Khalida Wani, Terri Armstrong, Elizabeth Vera-Bolanos, Aditya Raghunathan, David Ellison, Richard Gilbertson, Brian Vaillant, Stewart Goldman, Roger J. Packer, Maryam Fouladi, Ian Pollack, Tom Mikkelsen, Michael Prados, Antonio Omuro, Riccardo Soffietti, Alicia Ledoux, Charmaine Wilson, Lihong Long, Mark Gilbert, and Ken Aldape, for the Collaborative Ependymoma Research Network

Khalida Wani, Department of Pathology, The University of Texas M.D. Anderson Cancer Center, Houston, Texas 77030, USA; kaldape@mdanderson.org

ABSTRACT

Patients with ependymoma exhibit a wide range of clinical outcomes that is currently unexplained by clinical or histological factors. Little is known regarding molecular biomarkers that could predict clinical behavior. Since recent data suggests that these tumors display biological characteristics according to their location (cerebral vs. infratentorial vs. spinal cord), rather than explore a broad spectrum of ependymoma, we focused on molecular alterations in ependymomas arising in the infratentorial compartment. Unsupervised clustering of available gene expression microarray data revealed two major subgroups of infratentorial ependymoma. Group 1 tumors over expressed genes that were associated with mesenchyme, Group 2 tumors showed no distinct gene ontologies. To assess the prognostic significance of these gene expression subgroups, real-time reverse-transcriptase polymerase chain reaction assays were performed on genes defining the subgroups in a training set. This resulted in a 10-gene prognostic signature. Multivariate analysis showed that the 10-gene signature was an independent predictor of recurrence-free survival after adjusting for clinical factors. Evaluation of an external dataset describing subgroups of infratentorial ependymomas showed concordance of subgroup definition, including validation of the mesenchymal subclass. Importantly, the 10-gene signature was validated as a predictor of recurrence-free survival in this dataset. Taken together, the results indicate a link between clinical outcome and biologically-identified subsets of infratentorial ependymoma and offer the potential for prognostic testing to estimate clinical aggressiveness in these tumors.

Keywords: Infratentorial ependymoma, Expression profiling, Gene expression signature, Prognostic genes, Microarray, Biomarker

INTRODUCTION

Ependymomas are tumors of glial origin that morphologically resemble ependymal cells lining the ventricles of the central nervous system. These tumors may develop supratentorially, infratentorially as well as in the spinal cord [13]. Ependymoma accounts for 3–6% of all CNS tumors. The incidence is higher is children and young adults where it is the second most common malignant brain tumor. In adults the most common location is the spinal cord while in children these tumors tend to occur more commonly intracranially within the posterior fossa. Histologically ependymomas are classified into three major subtypes: myxopapillary ependymoma (WHO grade

I), ependymoma (WHO grade II) and anaplastic ependymoma (WHO grade III) [22]. Surgery represents the first line of treatment followed by radiotherapy and chemotherapy for disease control [26]. The 5-year progression-free survival ranges between 30% and 60% with incomplete surgical resection, anaplastic histology and younger age at diagnosis known to be associated with poor prognosis [23]. Whereas in children the 5 year progression free survival rate is 14% [31], in adult cases of infratentorial ependymoma the 5-year disease free survival rates as high as 70% [12] can be achieved. Gross total resection which is an important prognostic factor is not always achieved [14]. In children, adjuvant radiotherapy, another prognostic factor in disease control, often cannot be employed due to young age of the patients. Despite histological similarities, tumors within the same stage show diverse clinical behavior indicating inherent underlying differences. Insights regarding the genetic alterations found in ependymoma have recently been forthcoming [16; 35], however, little progress has been made in improving the survival rates of ependymomas, indicating the need for novel therapeutic approaches.

Several studies have shown that ependymomas arising in the infratentorial, supratentorial and spinal cord locations have distinct genetic signatures and must therefore be viewed as distinct tumor entities [15; 16; 19; 24]. Members of the Notch and Sonic Hedgehog pathway are highly expressed in intracranial ependymomas whereas the Homeo-box containing (HOX) family genes have been implicated in spinal cord ependymomas [35]. Data from one of our laboratories suggests that ependymomas derive from regionally specific stem cells bearing a radial glial cell phenotype [35]. Apart from histological grade and incomplete resection, several prognostic markers have been studied in ependymomas including Ki-67 [3], survivin [1], human telomerase reverse transcriptase, ERBB family members [11] and nucleolin [30]. Whole genome approaches such as array comparative genomic hybridization [20; 24; 28] and gene expression profiling [8; 16; 25; 27; 34] have also been employed for the identification of prognostic biomarkers. However, despite these studies molecular markers for disease prognosis have not been integrated into the clinical management of ependymomas. Availability of robust molecular markers of tumor recurrence would allow for the identification of those patients who would benefit from standard treatments. This would provide clinicians with the ability to identify those patients who would not respond to standard therapy thus allowing them to design more aggressive treatment strategies. Since ependymomas are largely chemo resistant, identification of molecular markers of prognosis would also help identify novel therapeutic targets

In this study, we examine biologic subtypes of infratentorial ependymoma identified by gene expression microarray data, as well as clinical implications of these subtypes. To achieve this, we examined gene expression profiling microarray data to identify robust biomarkers and applied them to clinically annotated routinely available formalin-fixed paraffin embedded (FFPE) samples. The biologic subtypes and the genes identified as associated with prognosis are validated in an external dataset.

MATERIALS AND METHODS

Tumor samples and patient characteristics

Archived paraffin embedded tissue specimens were collected from 56 patients with infratentorial ependymomas. All patients with tissue sufficient for gene expression assessment and a minimum follow up of 3 years or recurrence within 3 years were considered evaluable. For data analysis, cases were dichotomized into short recurrence-free survivors (SRS) and long recurrence-free survivors (LRS). LRS patients were defined as having clinical follow up data available for a minimum of 3 years without any recurrence within 3 years from date of surgery and SRS patients were defined as having had a recurrence within 3 years from the date of surgery. Histological diagnosis and tumor grading was performed based on WHO criteria [22]. Sections from paraffin embedded tissue were reviewed for pathologic diagnosis and dissected if necessary by a neuropathologist (KA) to ensure that ≥90% of the sample represented tumor. The clinical and pathological characteristics of the patients in the study are listed in Table 1. The use of the tissue and clinical data for these studies were covered under a protocol approved by the MD Anderson IRB.

Table 1

Clinical characteristics of the 56 infratentorial ependymoma patients.

Variable	Testing dataset n=18	Validation dataset n=38
Age		
< 18 yrs	12 (66.6%)	28 (73.7%)
> 18 yrs	6 (33.3%)	10 (26.3%)
WHO grade		
II	14 (77.8%)	32 (84.2%)
III	4 (22.2%)	6 (15.8%)
Extent of resection		
Gross total	15 (83.3%)	23 (60.5%)
Sub total	3 (16.6%)	13 (34.2%)
Unknown	0 (0%)	2 (5.3%)

Gene expression and array dataset

Gene expression data from four independent data sets from individual institutions was used for initial candidate biomarker discovery. Publicly available Affymetrix GeneChip data (.cel files) were obtained for data sets from the St Jude Children's Research Hospital, (GSE21687) [16], Virginia Commonwealth University (GSE13267 unpublished), University of Colorado-Denver (GSE16155) [8] and the Instituto Nazionale dei Tumori [24]. The Affymetrix Human Genome U133A v2.0 Array platform was used in three of the four datasets. Expression data was available for 67 infratentorial ependymomas from these sources. Microarray data were background corrected and

normalized using the guanine cytosine robust multi array average (gcRMA) algorithm in the Genespring software (Agilent biotechnologies) resulting in log2 expression values. For further validation, publicly available processed data files for Affymetrix Exon Array data from the Hospital for Sick Children, Toronto (GSE27279) and Agilent microarray data from the DKFZ, Heidelberg (GSE27287) were obtained [37].

Identification of candidate biomarkers

Within the sample sets, average expression and standard deviation values were calculated for each probe. The top 250 highly expressed highly variable probes were used for unsupervised clustering using Cluster 3.0 (http://bonsai.hgc.jp/~mdehoon/software/cluster) and Java tree view (http://jtreeview.sourceforge.net/) in order to identify subgroups based on gene expression patterns. Each sample was assigned to one of two identified gene expression subgroups. Fold change was utilized to identify genes whose expression differed between samples in each subgroup compared to tumors of the other subgroup. Genes were functionally annotated in DAVID (http://david.abcc.ncifcrf.gov/). Pathways altered in Group 1 and Group 2 ependymomas were identified using Gene Set Enrichment Analysis. The KEGG, Biocarta and GO databases were curated for pathway analysis. Candidate biomarkers for infratentorial ependymomas were identified based on high expression, high variability and fold change between the expression-defined subclasses. For candidate biomarker identification, average expression and differential fold change between Group 1 and Group 2 samples was calculated for each probe set. Among highly expressed genes (top 50th percentile of mean expression), genes with the largest fold change between subclass were identified. 295 were selected for further study. Additional genes (n=71) that were defined in a prior publication on ependymoma subclasses [16] were also included.

RNA extraction

Total cellular RNA was isolated from core punches using the Epicenter RNA isolation kit according to the manufacturer's protocol (Epicenter Biotechnologies, Madison, WI) following de-paraffinization and proteinase K treatment. Genomic DNA contamination was removed through a DNAse treatment step in the protocol. Quality of the RNA was assessed using a real-time PCR assay for the control gene EEF1A (Assay Id: Hs00265885_g1; Applied Biosystems, Foster city, CA, USA) using 1μl of undiluted cDNA and standard TaqMan cycling conditions on a 7900HT Real-Time PCR machine (Applied Biosystems). Samples with a Ct value of ≤28 were considered eligible for gene expression analysis.

Real-time PCR

A custom real-time PCR based micro fluidics card (Applied Biosystems Foster City, CA) was designed and populated with 366 candidate biomarker genes identified in the discovery microarray expression dataset. In order to optimize amplification of the fragmented RNA found in FFPE processed tissue, TaqMan gene expression primers were picked with predicted amplicon sizes of 85

base pairs or less (Applied Biosystems, Foster City, CA) (Supplementary Table 1). Quantitative real-time PCR (QRT-PCR) measurements were performed using a set of 18 FFPE ependymomas (9 SRS and 9 LRS). Total tumor RNA was reverse transcribed to single-stranded cDNA using ABI's high capacity reverse transcription kit using the maximum allowed concentration of total RNA per manufacturer's instructions (100ng/μl). To determine fold-changes in each gene, QRT-PCR was performed. Fold-change associated with survival for each gene was determined by the ΔΔCt method.

Since FFPE-derived RNA can be highly degraded, some quality metrics were employed. The quality metric was to determine the ability to amplify a housekeeping gene (EEF1A). This measurement was then used to guide the RT-PCR for the mRNA microfluidics gene card. Second, since gene assays with very high Ct values can mean either low expression or poor quality RNA, we examined the Ct values of the control genes GLUD2, NTF3 and PPIC, which were selected for high expression and relatively uniform expression across ependymoma samples in the discovery microarray datasets. If the average Ct value of the 3 control genes was over 34, the sample was considered to have RNA quality too low for a reliable measurement. In order to prioritize highly expressed genes which are more reliably measured in FFPE tissues, we used an average delta Ct cutoff of 4 cycles or less for a gene to be considered for further testing. This occurred in 55% of gene assays among all the samples. Candidate biomarkers from this subset that were found to be associated with recurrence-free survival were identified and included in a validation test. Validation was performed on an independent set of 38 ependymomas (17 LRS and 21 SRS). Genes associated with survival were identified based on fold change (1.5 fold and higher) and t test (p ≤ 0.05).

Calculation of metagene score

In order to determine the association of the overall gene expression classifier with patient outcome, we calculated a single "metagene" score as described in our previous publication [6] with the exception that all the genes were weighted equally. For each case the metagene score was calculated based on the set of 10 genes by averaging the normalized expression values for all the genes associated with poor prognosis and then subtracting the average of the normalized expression values for all the genes associated with good prognosis for each case. This resulted in a single numerical score for each tumor, and each tumor was then ranked according to this metagene score. The samples were dichotomized into 2 groups using median metagene score as the cutoff. Samples with metagene scores above the median were defined as having a favorable 10-gene signature while samples with metagene scores below the median value were defined as having an unfavorable 10-gene signature.

Immunohistochemistry

Immunohistochemical staining on a set of 39 archival FFPE infratentorial ependymoma tumors for TOP2A was carried out using a mouse monoclonal antibody (clone Ki-S1; Dako; 1:100). Antigen retrieval was done using heat-induced antigen retrieval in 10mM sodium citrate buffer, pH 6.0.

Staining was detected using the Envision kit from Dako. Expression of TOP2A protein was scored as present (1) or absent (0).

Statistical analysis

All data were analyzed with Statistica 6.1 software package (Statsoft Inc). Univariate and multivariate Cox regression analysis evaluated the association of individual and multiple covariates with recurrence-free survival. Kaplan – Meier estimated the time to event functions of recurrence-free and overall survival. Recurrence-free survival was defined as the time between date of surgery to date of recurrence, death or last follow up. A $p \leq 0.05$ was considered significant.

RESULTS

Unsupervised clustering to identify transcriptomal groups

The overall experimental scheme is outlined in Supplementary Fig. 1. We first attempted to identify transcriptomal subtypes of infratentorial ependymoma, by obtaining publicly available Affymetrix U133 profiling data from 4 sources (including data from one of our laboratories) comprising samples from 67 patients. After normalization, unsupervised clustering was performed on genes that were both highly expressed and highly variable in the expression array data. Unsupervised clustering revealed 2 major subclasses of these tumors, as shown in Fig. 1. Patient age was available for 59/67 cases (Fig. 2). Distribution of the patient samples showed that Group 1was composed predominantly of children. Of the 24 patients for whom age at diagnosis data was available, 23 were under the age of 10 years and only 1 child was over 10 years. In contrast, 7/30 samples from Group 2 were from patients over the age of 10, which included 4 adults (age > 18) (chi-square p = 0.025). Our findings have been further validated in the Witt et al publication wherein the authors have found a strong correlation between younger patient age and a molecular subgroup characterized by the up regulation of pathways known to be activated in aggressive tumors. Gene ontology (GO) analysis showed that genes over expressed in Group 1 samples had ontologies correlating with response to wound healing, inflammation, migration and cell adhesion (Table 2 and Supplementary Table 2). These gene ontologies are similar to previous findings from our laboratory and others [10; 29; 36] which described a mesenchymal signature in glioblastoma (GBM). To examine this further, we evaluated the expression of genes that we previously described as mesenchymal [29]. Of the 15 genes originally described in the manuscript, 13 were interrogated on the microarray platform used for these data (U133A GeneChip). All 13 of these genes were over expressed in Group 1 compared to Group 2 (Supplementary Table 3), indicating concordance of the Group 1 signature with mesenchymal glioblastoma. No significant GO terms were identified for genes over expressed in Group 2 (Supplementary Table 4). Comparison of infratentorial subgroups (Subgroups G, H and I) described in a prior publication [16] showed that the Group 1 is essentially similar to subgroups H and I while Group 2 resembles subgroup G (Supplementary Fig. 2).
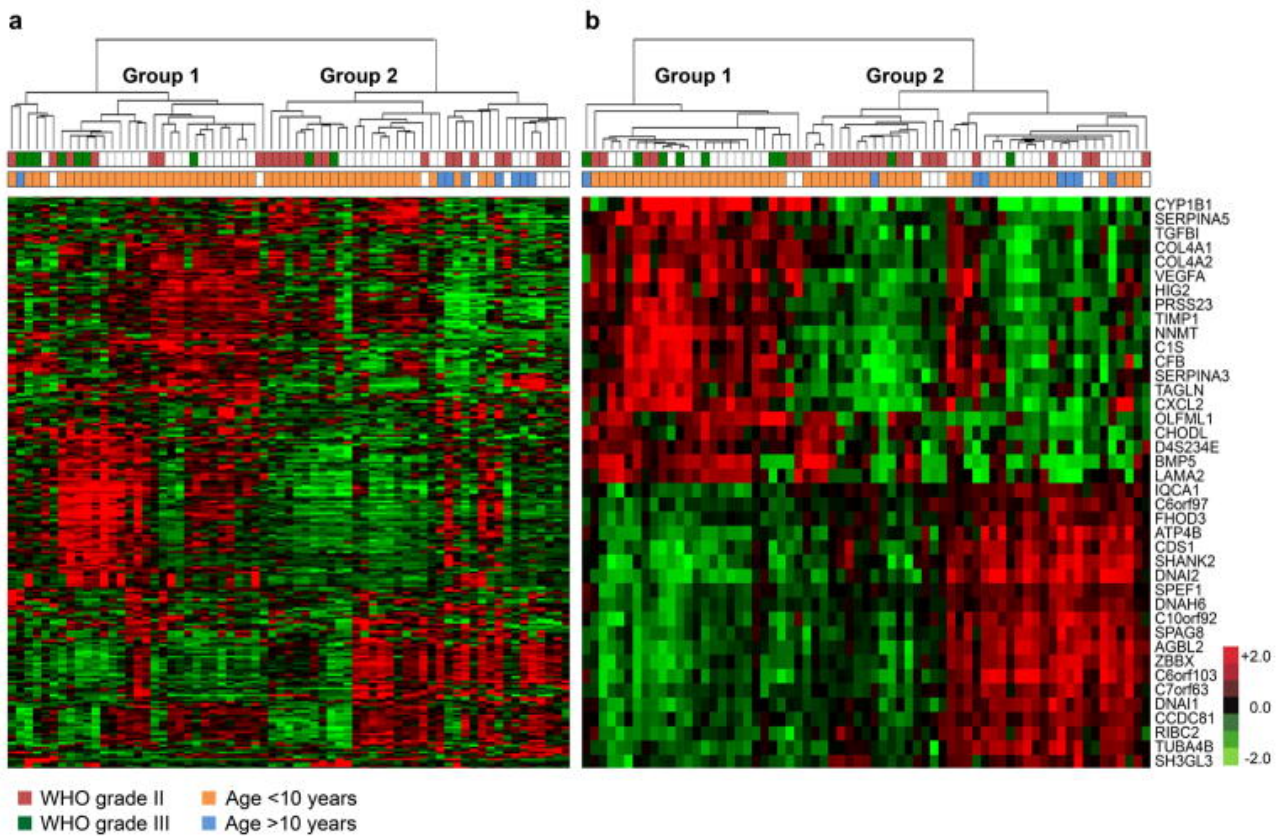
Fig.1 Expression profiling reveals two major subgroups of gene expression in infratentorial ependymomas. **a)** Unsupervised clustering of 67 infratentorial ependymomas using 250 highly variable probe sets reveals two major sample clusters. Age: patients <10 years (orange), >10 years (blue) Grade: II (red), III (green) Unknown (white). **b)** Heat map of the top 40 genes differentially expressed between group 1 and group 2.

Table 2

Gene ontology analysis of genes over expressed in microarray expression Group 1.

GO TERM ID	Function	Count	Benjamini
9611	Response to wounding	45	< 0.0005
6954	Inflammatory response	31	< 0.0005
6952	Defense response	38	< 0.0005
6955	Immune response	40	< 0.0005
10033	Response to organic substance	37	< 0.0005
7626	Locomotory behavior	21	< 0.0005
1501	Skeletal system development	23	< 0.0005
42127	Regulation of cell proliferation	38	< 0.0005
22610	Biological adhesion	34	< 0.0005
7155	Cell adhesion	34	< 0.0005
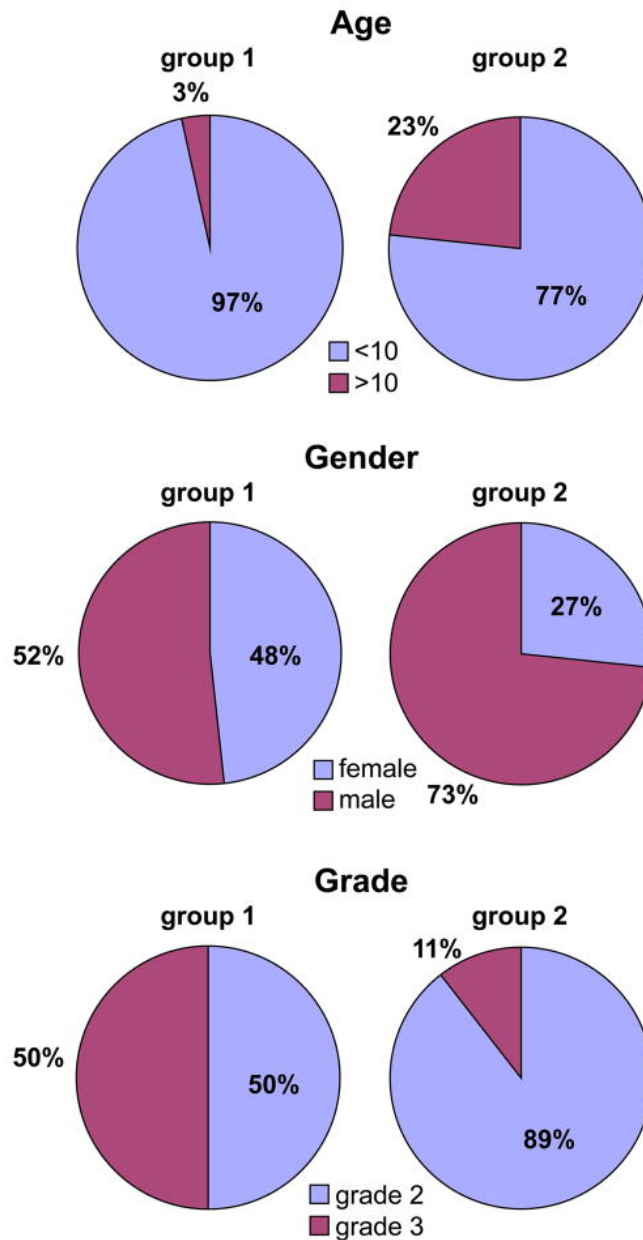
Fig. 2 Distribution of age, grade and gender among the microarray defined subgroups. Group 1 ependymomas were predominantly composed of younger patients.

Identification of prognostic gene set

Genes that define the mesenchymal and proneural subtypes of GBM have been found to function as robust biomarkers of clinical outcome [6]. Since several of the genes over expressed in the mesenchymal glioblastoma are also over expressed in Group 1 ependymomas we reasoned that some of the genes that defined transcriptomal subtypes might also be associated with patient outcome in ependymoma. To test this, we selected 366 genes that were differentially expressed between the 2 groups and conducted quantitative real-time PCR for these genes on a set of 18 formalin-fixed, paraffin embedded (FFPE) archival infratentorial ependymoma samples on which clinical data were available (Supplementary Table 5). Using 3-year recurrence-free survival as a

cutoff, we identified a set of 51 genes (Supplementary Table 6) that were differentially expressed between patients with short recurrence-free survival (SRS) and long recurrence-free survival (LRS). Of these 51 genes, 22 were found to be over expressed in the SRS patient samples and the remaining 29 were over expressed in the LRS patient samples. Comparison with unsupervised clustering showed a striking correlation with transcriptomal subgroups. Twenty of the 22 SRS-associated genes were over expressed in cluster Group 1 tumors, while twenty seven of the 29 LRS-associated genes were over expressed in cluster Group 2 (Supplementary Table 6). These results suggest a significant concordance of outcome-associated genes with genes that define global transcriptomal classes (Fischer's exact test, 2-tailed p < 0.001) (Supplementary Table 7).

We then tested these 51 genes in an independent sample of 38 infratentorial ependymoma samples (21 SRS and 17 LRS). Using a metagene approach we found that in univariate Cox analysis, the 51-gene signature was associated with recurrence-free survival (p = 0.016) as well as overall survival (p = 0.0068). In order to optimize this signature and increase its applicability to routine clinical testing we focused on the top differentially expressed genes between the SRS and LRS group of samples. To accomplish this, we used a fold-change cutoff of 1.8 or higher and a t-test $p \leq 0.05$, which resulted in a 10-gene signature (Table 3). To determine how the 10-gene signature compares with current standard clinical parameters as a prognostic factor, univariate Cox regression analyses were performed on data from all the 56 samples. In univariate analyses, the 10-gene signature was strongly associated with recurrence-free survival (p < 0.001) as well as overall survival (p < 0.001) (Table 4). Using a cutoff of median metagene score, Kaplan-Meier analyses demonstrate a significant difference in both recurrence-free and overall survival based on the 10-gene score (Fig. 3). The median recurrence-free survival for the metagene defined unfavorable patient group was 80 weeks while the median for the favorable group was not reached. Consistent with the findings from microarray data, the unfavorable 10-gene signature was associated with a younger age group. Of the 28 patients with an unfavorable 10-gene signature, 22 were below the age of 10. In contrast, only 9/28 patients with the favorable signature were under 10 (Fischer's exact test, 2-tailed p = 0.001) (Fig. 4).

Table 3

The survival-associated 10-gene expression signature of infratentorial ependymoma.

| Gene	Assay ID	Fold change (SRS vs. LRS)	p-value[*]	Transcriptomal grou
MMP9	Hs00234579_m1	7.04	< 0.001	Group 1
TOP2A	Hs03063307_m1	6.62	< 0.001	Group 1
TKTL1	Hs00202061_m1	5.02	0.003	Group 1
COL3A1	Hs00943809_m1	4.48	< 0.001	Group 1
LAMB1	Hs01055971_m1	3.45	0.001	Group 1
COL4A2	Hs01098873_m1	3.00	0.002	Group 1
TGFBI	Hs00932747_m1	2.22	0.003	Group 1
GRIA1	Hs00181348_m1	0.28	< 0.001	Group 2
F5	Hs00914120_m1	0.23	< 0.001	Group 2
SLC14A1	Hs00998197_m1	0.16	0.003	Group 2
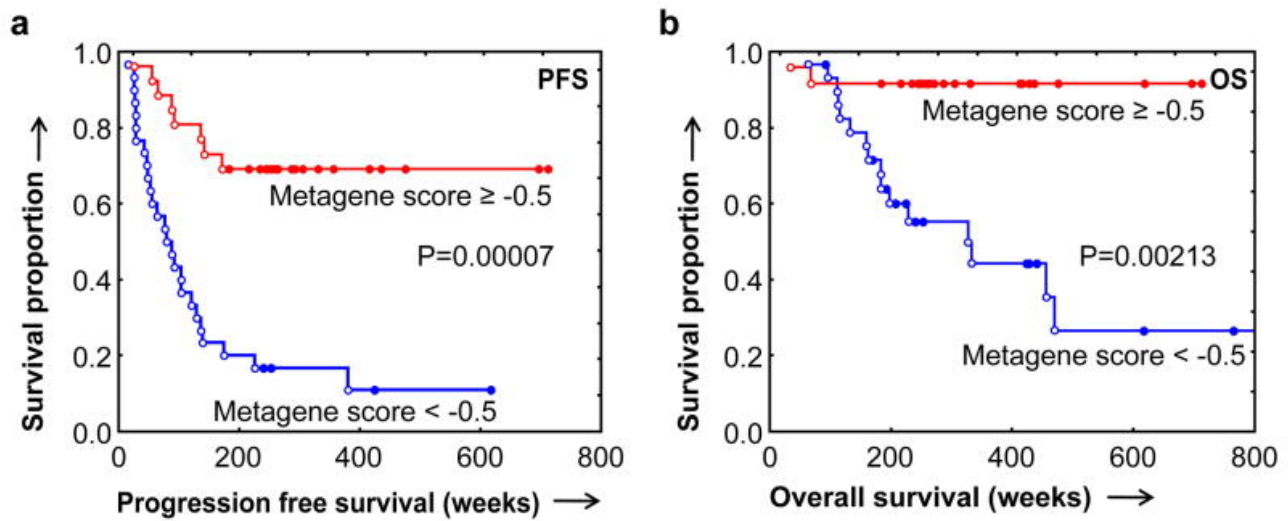
Fig. 3 Kaplan-Meier analysis of survival association with 10-gene signature. Formalin-fixed, paraffin embedded ependymoma samples were subjected to QRT-PCR as described in Materials and Methods. The samples were ranked by metagene score and dichotomized into 2 groups using the median metagene score as the cutoff. Survival is shown for the lower metagene scores (blue) vs. the higher metagene score (red). **a)** Recurrence-free survival according to the 10-gene set. The median recurrence-free survival for the metagene-defined unfavorable group was 80 weeks while for the favorable group it was not reached. **b)** Overall survival according to the 10-gene set. The median overall survival for the unfavorable group was 345 weeks, and was not reached for the favorable group. The log rank test was used to determine statistical significance.

Table 4

Univariate analysis of outcome predictors.

Variable	Recurrence-free survival		Overall survival	
	HR	p-value	HR	p-value
Age	0.98	0.096	0.96	0.040
Extent of resection	1.60	0.187	1.49	0.406
WHO Grade	5.10	< 0.001	8.40	< 0.001
Adjuvant XRT	0.37	0.013	0.25	0.028
Metagene score	0.80	< 0.001	0.74	< 0.001

Gross total resection = 1, Subtotal resection = 0; Adjuvant radiotherapy: yes = 1, no = 0 continuous variable.
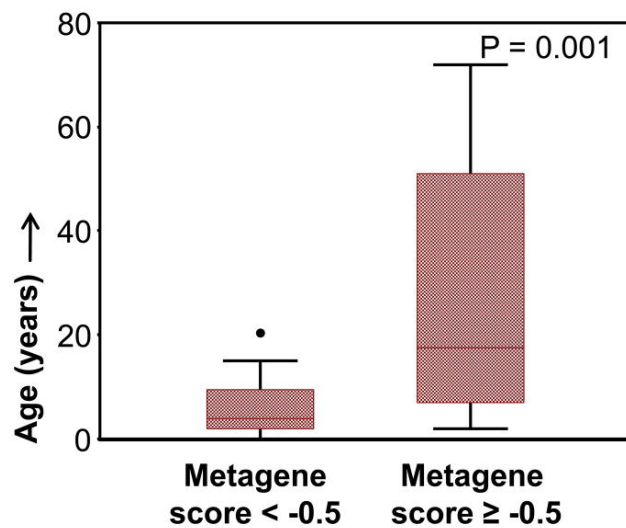
Fig. 4 The unfavorable 10-gene signature is significantly associated with younger age at diagnosis. Statistical significance of age was determined using the Fisher's exact test, two tailed.

Multivariate analyses were then performed to determine whether the 10-gene signature was independently associated with survival outcomes in the context of prognostic factors found to be significant in univariate analysis. In multivariate analysis, the 10-gene signature was significantly associated with recurrence-free survival (p = 0.003) (Table 5). One of the genes in the 10-marker set, Topoisomerase 2 alpha (TOP2A), was tested using immunohistochemistry (IHC) on a set of 39 infratentorial ependymomas. TOP2A expression was significantly associated with the metagene score (Fishers exact test, 2-tailed p < 0.001) (Supplementary Fig. 3). When grouped based on expression of TOP2A, Kaplan-Meier analyses demonstrate a significant difference in both recurrence-free and overall survival among groups with low and high expression of TOP2A (Supplementary Fig. 3). In the Heidelberg dataset however, expression of TOP2A mRNA was not significantly associated with recurrence-free survival (log rank test p = 0.12) or overall survival (log rank test p = 0.08) when grouped by median gene expression.

Table 5

Multivariate Cox Regression analysis.

Variable	Recurrence-free survival		Overall survival	
	HR	p-value	HR	p-value
Age	0.99	0.601	0.98	0.232
Extent of resection	2.66	0.014	2.45	0.075
WHO Grade	0.97	0.944	3.14	0.124
Adjuvant XRT	0.52	0.133	0.40	0.150
Metagene score	0.79	0.003	0.84	0.104

Gross total resection = 1, Subtotal resection = 0; Adjuvant radiotherapy: yes = 1, no = 0 continuous variable.

**Validation of microarray defined subgroups**

Following primary analysis of our data an independent group reported results from mRNA-based expression profiles in infratentorial ependymoma [37]. Briefly, 2 major subtypes (Groups A and B) were identified in this report, with biologic and clinical implications. We obtained the gene expression data used in this paper (GSE27287, GSE27279) to validate our findings. The paper describes 2 datasets consisting of a total of 84 infratentorial samples, the first tested with an Affymetrix exon array platform (n=47, "Toronto" dataset) and the second with an Agilent platform (n=37, "Heidelberg" dataset). We clustered samples from these datasets using the gene list that defined our Groups 1 and 2 subtypes (Fig 2B). This resulted in two major subgroups in the 2 datasets (Fig. 5), showing significant concordance in subgroup assignment (Table 6). Conversely, we applied the independently defined gene list to cluster the microarray data from the 67 samples described above, which again resulted in a significant concordance in subgroup assignment. (Table 7). In the Heidelberg dataset clustered using the gene list that discriminates Group 1 and 2, molecular subclass was significantly associated with recurrence free survival ($p = 0.041$) as well as overall survival ($p = 0.015$) in univariate analysis (Supplementary Table 8). Finally, we examined the 2 subgroup-defining genes indicated in the combined Toronto-Heidelberg datasets LAMA2 and NELL2 in our dataset. The Group A marker LAMA2 was highly expressed in microarray Group 1 (Student's t-test, $p < 0.001$) while the Group B marker NELL2 was over expressed in microarray Group 2 (Student's t-test, $p < 0.001$) (Fig. 6). Analysis of the pathways defined by gene ontologies in Group 1 versus Group A in the Toronto and Heidelberg datasets by GSEA showed 32 commonly deregulated pathways. Among Group 2 and Group B from the two datasets there were 18 commonly deregulated pathways. By random chance less than 1 pathway would be expected to be commonly altered between Group 1/Group A and Group 2/Group B providing further evidence that the subgroups in our analysis display similar biology to subgroups recently described in an independent dataset.
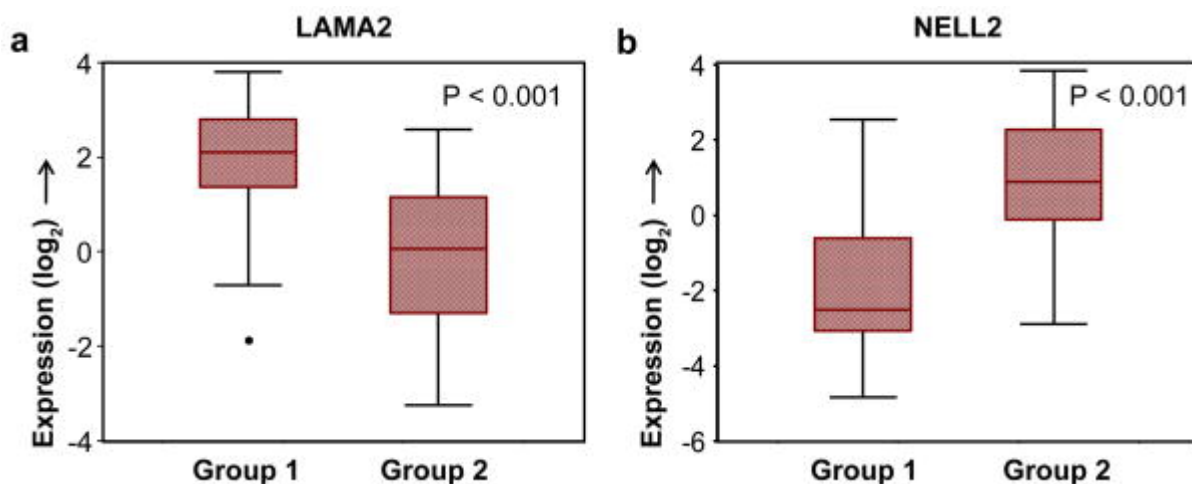
Table 6

Concordance between the microarray-defined subgroups: clustering of external datase gene signature described in Figure2B.

	Toronto dataset				Heidelberg dataset		
	Group A	Group B	p-value[*]		Group A	Group B	p-value[*]
Group 1	29	0	< 0.001	Group 1	13	0	< 0.001
Group 2	5	13		Group 2	5	19	

[*]Fisher's exact test two-tailed
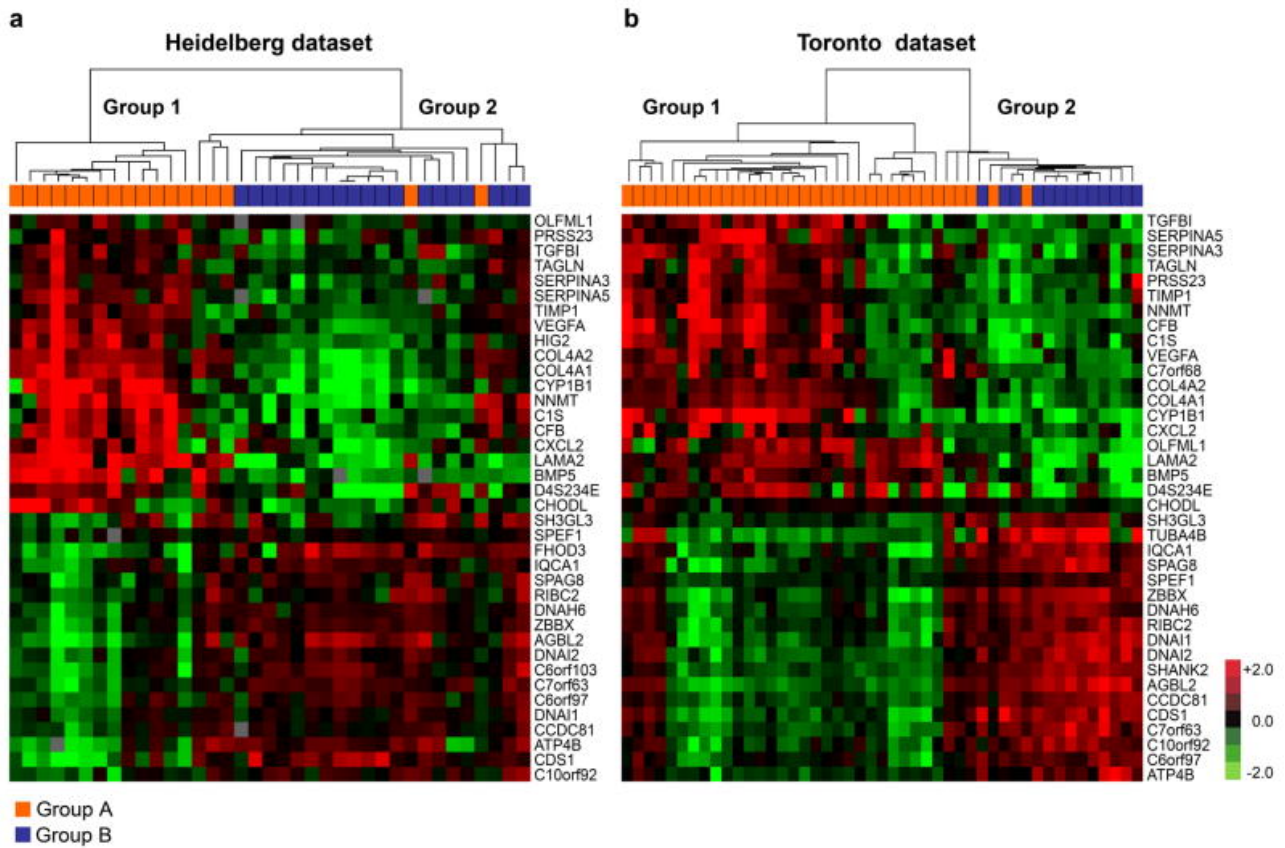
Fig. 5 Subgroups A and B described in the Witt et al report are similar to transcriptomal subgroups 1 and 2. Gene expression microarray data from the Toronto and Heidelberg datasets was used for clustering using the gene list described in Fig. 2B. Heat map of the top 40 genes differentially expressed between Group 1 and Group 2 in the a) Heidelberg dataset and b) the Toronto dataset. 38 of the 40 most differentially expressed genes in Groups 1 and 2 were evaluable in the Heidelberg data set while 38 of the 40 differential genes were present in the Toronto dataset. Color coding is as follows, Group A tumors: orange; Group B tumors: blue.


Fig. 6 Expression of LAMA2 and NELL2 in the transcriptomal subgroups. LAMA2 was over expressed in Group 1 ependymomas while NELL2 expression was significantly higher in Group 2 ependymomas. The Student's t-test was used to determine statistical significance.

Table 7 Concordance between the microarray-defined subgroups: clustering of samples described in Figure 2B using gene signature described in Witt et al.

	Group A	Group B	p-value*
Group 1	25	1	
			0.003
Group 2	13	28	

*Fishers exact test two-tailed

Finally, we applied the 10-gene signature to the external dataset to validate outcome association. Univariate analyses validated associations with recurrence-free survival and overall survival (Table 8 and Fig. 7). In multivariate analysis, the 10-gene signature was validated as a predictor of recurrence-free survival in the independent dataset (Table 9).
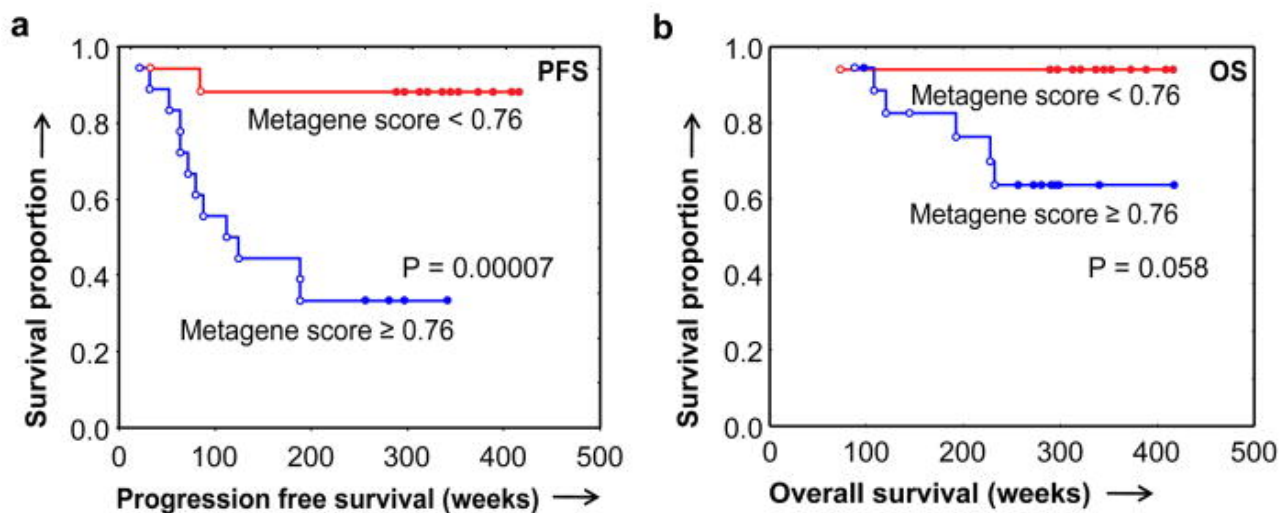
Fig. 7 Kaplan-Meier analysis of survival association with 10-gene signature in the Heidelberg dataset. Gene expression data from the Heidelberg dataset was used to calculate a metagene score for the 10-gene signature as described in materials and methods. The samples were ranked by metagene score and dichotomized into 2 groups using the median metagene score as the cutoff. Survival is shown for the higher metagene scores (blue) vs. the lower metagene score (red). **a)** Recurrence-free survival according to the 10-gene set. The median recurrence-free survival for the metagene-defined unfavorable group was 110 weeks while for the favorable group it was not reached. **b)** Overall survival according to the 10-gene set. The log rank test was used to determine statistical significance

Table 8

Univariate analysis of outcome predictors in the Heidelberg dataset.

Variable	Recurrence-free survival		Overall survival	
	HR	p-value	HR	p-value
Age	0.96	0.108	0.94	0.136
Extent of resection	0.11	< 0.001	0.09	0.024
WHO Grade	3.56	0.052	5.30	0.123
Adjuvant XRT	1.76	0.338	0.88	0.861
Metagene score	1.91	< 0.001	2.34	0.002

Gross total resection = 1, Subtotal resection = 0; Adjuvant radiotherapy: yes = 1, no = (
continuous variable.

Table 9

Multivariate Cox Regression analysis of the Heidelberg dataset.

Variable	Recurrence-free survival		Overall survival	
	HR	p-value	HR	p-value
Age	0.94	0.133	0.98	0.688
Extent of resection	0.02	< 0.001	0.10	0.042
WHO Grade	8.21	0.023	5.11	0.218
Adjuvant XRT	11.78	0.016	1.23	0.850
Metagene score	1.71	0.026	1.92	0.100

Gross total resection = 1, Subtotal resection= 0 ; Adjuvant radiotherapy: yes = 1, no = (
continuous variable.

DISCUSSION

Despite histological similarities, ependymomas arising from the spinal cord and the infratentorial and supratentorial compartments of the central nervous system show diverse clinical behavior [2]. Several gene expression profiling studies have shown that ependymomas of different locations indeed have distinct expression profiles suggesting biological tumor heterogeneity [9; 15; 16; 24]. To address site-specific biological variation, we focused on infratentorial ependymomas, one of the common locations found in adult and pediatric tumors. Since infratentorial ependymomas have distinct gene expression signatures, we hypothesized that they would have a unique set of biologic and prognostic markers relative to ependymomas arising in other sites.

Using unsupervised clustering of publicly available infratentorial ependymoma microarray expression data, we found two distinct molecular subtypes of infratentorial ependymoma characterized by distinct biological pathways. Gene ontology analysis showed that one of the subgroups identified in unsupervised clustering (Group 1) showed over expression of genes associated with angiogenesis, wound healing, migration and adhesion. These gene ontologies are similar to the mesenchymal gene signature reported in glioblastoma [29; 36]. In our analysis of genes that we previously described as prototypical mesenchymal genes, we found that these genes were uniformly over expressed in Group 1 ependymomas. Patients within expression Group 1 were on average, younger than patients whose tumors clustered as Group 2. While the number of adult patients in the microarray analysis was small (n=4), all 4 patient samples clustered within Group 2. All of the patients in Group 1 were children. These results may suggest age-specific biologic differences in infratentorial ependymoma. This conclusion is supported by prior studies that have highlighted genetic [19], epigenetic [32] and chromosomal characteristics [24] of pediatric ependymomas that differentiate them from the molecular profile of adult tumors. A recent report on infratentorial ependymoma expression profiling wherein two molecularly distinct subgroups were identified [37], validates our findings. This report describes a subgroup of infratentorial ependymoma that is comprised of younger patients and defines a more aggressive phenotype. We have shown that these subgroups are essentially similar to the ones we have described, and describe similar biology, despite differences in the individual genes that define them.

Several groups have reported the prognostic value of gene expression signatures identified by global expression profiling [4; 10; 29; 33]. In order to explore the clinical significance of the microarray-defined subclasses, we tested a panel of genes that were differentially expressed in the two expression groups on a set of clinically annotated infratentorial ependymomas. Using a small pilot sample (n=18) we identified a set of 51 genes that showed an association with recurrence- free as well as overall survival. This 51-gene signature remained prognostic in a larger independent validation set. Correlation of the 51 gene signature with the microarray defined classes showed that most (approximately 90%) of the genes over expressed in the tumors that recurred (SRS) clustered within the microarray expression Group 1, which represented a more aggressive molecular profile. In contrast, the vast majority of genes highly expressed in the LRS tumors were over expressed in microarray Group 2. These findings indicate close connections between microarray-defined subgroups, patient age and prognostic biomarkers in infratentorial ependymoma.

In order to increase the feasibility of the development of a clinical test applicable to routine testing, we selected the 10 genes most significantly associated with survival. Expression values of these 10-

gene assays were condensed into a single score (metagene) for statistical analyses. The 10-gene set was shown to be an independent predictor of recurrence-free survival as well as overall survival. While there was a strong association between younger age and an unfavorable prognostic signature, the 10-gene signature remained significantly associated with patient outcome after adjusting for age. We have validated the 10-gene signature as an independent predictor of recurrence-free survival as well as overall survival in the Heidelberg dataset as recently described [37]. Importantly, while the larger 51-gene signature and the more focused 10-gene signature were optimized using clinical outcome as the endpoint, these genes were derived from a set of genes differentially expressed between transcriptomic subgroups independent of patient outcome, indicating a connection between biologic subtypes of ependymoma and tumor aggressiveness (Supplementary Tables 6 and 7).

Analysis of the known functions of the genes in our 10-gene signature showed that poor survival is associated with the increased expression of genes associated with angiogenesis, extra cellular matrix and proliferation. Increased expression of the oncogene Transketolase-like 1 (TKTL1), has been reported to be associated with poor survival in colorectal carcinomas [7]. Over expression of TOP2A mRNA was associated with shorter metastasis-free survival in breast cancer, indicating that it could be a marker for aggressive tumor phenotype [5]. As seen in our study, overexpression of TOP2A has been reported to be associated with poor survival in ependymomas in univariate analysis [17; 18; 21; 38; 39]. However, TOP2A over expression has not been shown to represent an independent prognostic marker after accounting for clinical variables such as extent of resection and grade thus limiting its applicability as a prognostic biomarker.

These data provide preliminary evidence that an mRNA-based test could serve as a format for a clinical test for infratentorial ependymoma which, along with the existing clinical markers, could be used to optimize therapeutic choices for individual patients, analogous to the predictive test developed for optimization of patients therapy in glioblastoma [6] and breast cancer [33]. Specifically, with further validation, such a test could help identify the likelihood of response to standard therapies. In addition, the identification of specific clinical outcome-associated genes could also provide insights into tumor biology that could help identify novel therapies for patients resistant to standard therapy. Further validation of this 10-gene signature in a larger independent cohort of samples would be required to demonstrate the clinical utility of this molecular signature.
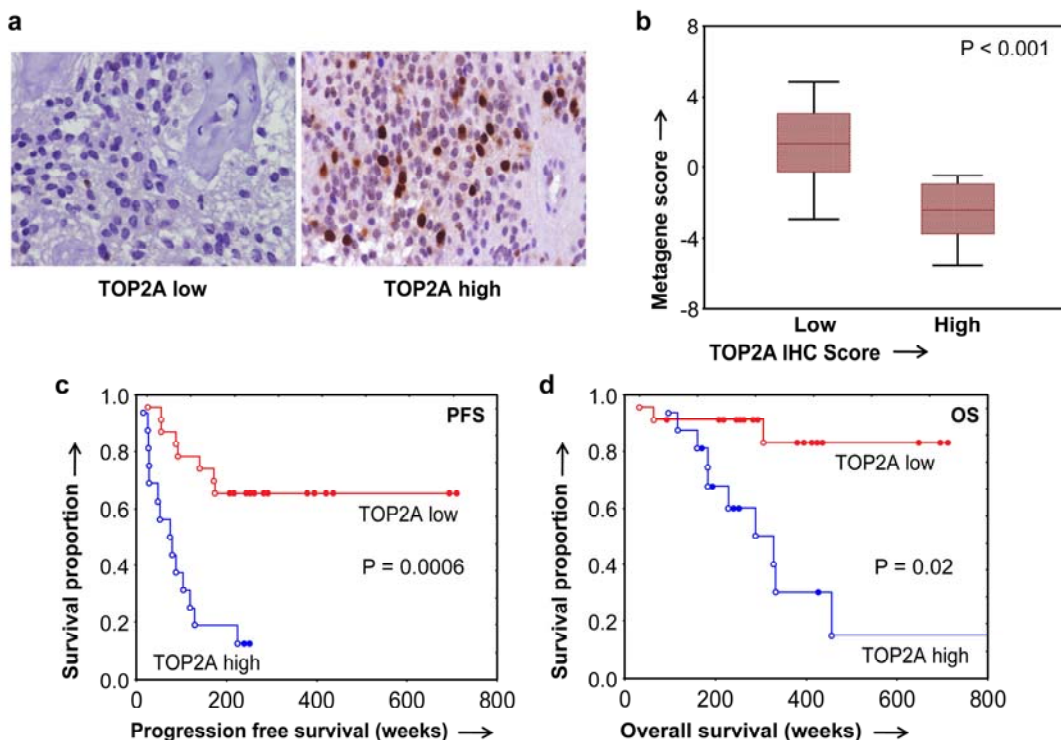
SUPPLEMENTARY MATERIAL

Supll. Fig. 1: Overall experimental approach for the study. The number of microarray samples from each site is indicated. INDT: Instituto Nazionale Dei Tumori; UCD: University of Colorado at Denver; SJCRH: St. Jude Children's Research Hospital; VCU: Virginia Commonwealth University.
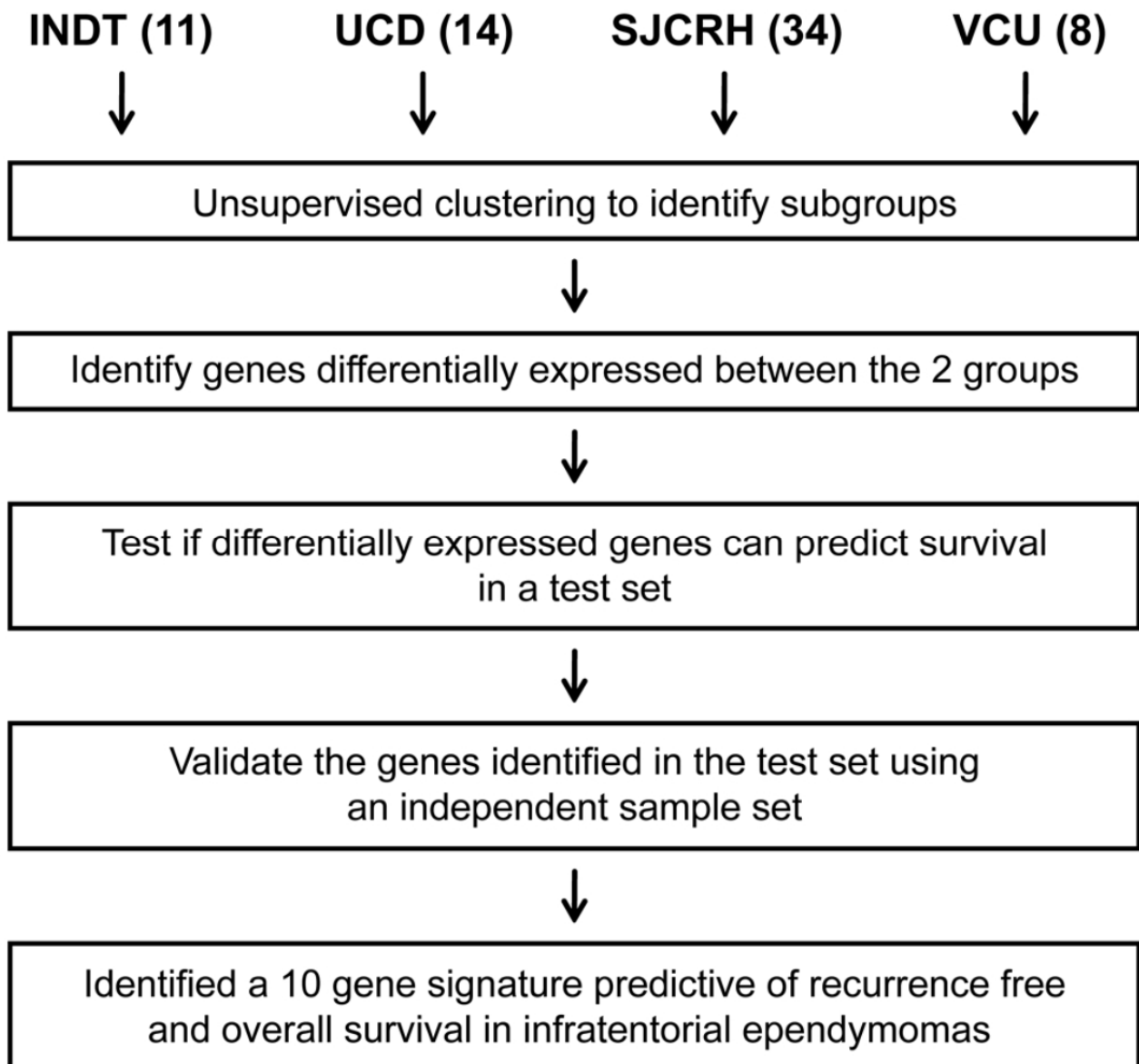
Supplementary Fig. 2: Comparison of the transcriptomal Groups 1 and 2 with the Johnson et al posterior fossa subgroups G, H and I. Group 1 tumors are similar to subgroup H and I tumors in their gene expression profile while group 2 tumors resemble subgroup G.
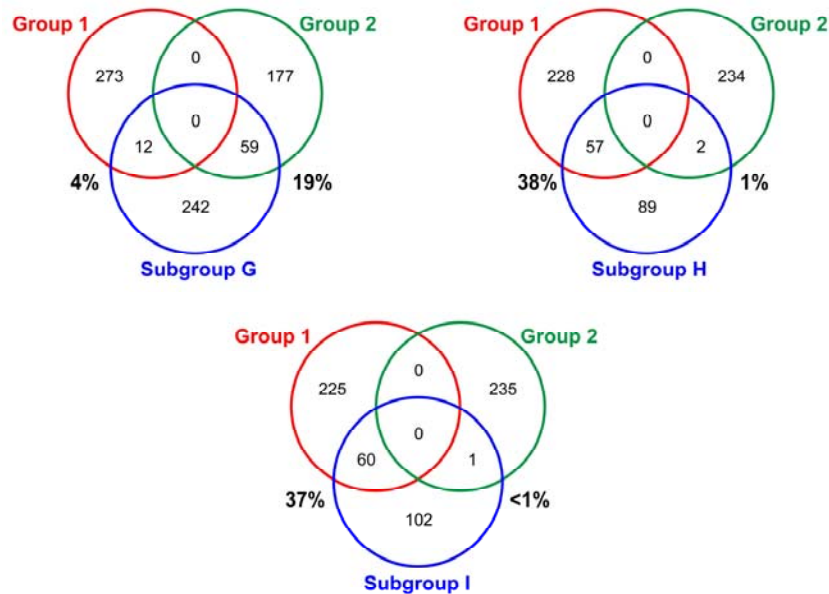
Supplementary Fig. 3 Expression of TOP2A and its correlation with metagene score and patient survival. a) Immunohistochemical staining for TOP2A showing a negative and positive case. b) Box-whisker plot showing the association of TOP2A staining and metagene score. High TOP2A immunostaining score corresponded with a low metagene score (unfavorable response group) while low TOP2A immunostaining was seen in the high metagene score group (favorable group).The Fisher's exact test, two-tailed, was used for statistical significance calculation. c-d) Kaplan-Meier survival curves showing a worse progression-free and overall survival in the group of cases that had high levels of TOP2A protein expression as detected by immunohistochemistry.

ACKNOWLEDGMENTS

Contributor Information

Khalida Wani, Department of Pathology, The University of Texas M.D. Anderson Cancer Center, Houston, Texas 77030, USA.

Terri Armstrong, Department of Family Health, The University of Texas Health Science Centre School of Nursing and Department of Neuro-Oncology, The University of Texas MD Anderson Cancer Centre, Houston, Texas, USA.

Elizabeth Vera-Bolanos, Department of Neuro-Oncology, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA.

Aditya Raghunathan, Department of Pathology, The University of Texas M.D. Anderson Cancer Center, Houston, Texas 77030, USA.

David Ellison, Department of Pathology, St Jude Children's Research Hospital, Memphis, Tennessee, USA.

Richard Gilbertson, Department of Developmental Neurobiology, St Jude Children's Research Hospital, Memphis, Tennessee, USA.

Brian Vaillant, Department of Neurosurgery, Methodist Neurological Institute, Houston, Texas, USA.

Stewart Goldman, Department of Neuro-Oncology, Children's Memorial Hospital, Chicago, IL, USA.

Roger J. Packer, Center for Neuroscience and Behavioral Medicine, Brain Tumor Institute, Children's National Medical Center for Neuroscience Research (CNR) Washington, DC, USA.

Maryam Fouladi, Department of Neuro-Oncology, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, USA.

Ian Pollack, Department of Neurological surgery, Children's Hospital of Pittsburgh, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA.

Tom Mikkelsen, Department of Neurosurgery, Henry Ford West Bloomfield Hospital, West Bloomfield, MI, USA.

Michael Prados, Department of Neurosurgery, University of California, San Francisco, San Francisco, CA, USA.

Antonio Omuro, Department of Neuro-Oncology, Memorial Sloan-Kettering Cancer Center, New York, NY, USA.

Riccardo Soffietti, Department of Neuro-Oncology, University of Torino, Medical School, Italy.

Alicia Ledoux, Department of Pathology, The University of Texas M.D. Anderson Cancer Center, Houston, Texas 77030, USA.

Charmaine Wilson, Department of Pathology, The University of Texas M.D. Anderson Cancer Center, Houston, Texas 77030, USA.

Lihong Long, Department of Pathology, The University of Texas M.D. Anderson Cancer Center, Houston, Texas 77030, USA.

Mark Gilbert, Department of Neuro-Oncology, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA.

Ken Aldape, Department of Pathology, The University of Texas M.D. Anderson Cancer Center, Houston, Texas 77030, USA.

REFERENCES

1. Altura RA, Olshefski RS, Jiang Y, Boue DR. Nuclear expression of Survivin in paediatric ependymomas and choroid plexus tumours correlates with morphologic tumour grade. *Br J Cancer*. 2003;89:1743–1749.

2. Armstrong TS, Vera-Bolanos E, Gilbert MR. Clinical course of adult patients with ependymoma: Results of the Adult Ependymoma Outcomes Project. *Cancer*. 2011;117:5133–5141.

3. Bennetto L, Foreman N, Harding B, Hayward R, Ironside J, Love S, Ellison D. Ki-67 immunolabelling index is a prognostic indicator in childhood posterior fossa ependymomas. *Neuropathol Appl Neurobiol*. 1998;24:434–440.

4. Brannon AR, Reddy A, Seiler M, et al. Molecular Stratification of Clear Cell Renal Cell Carcinoma by Consensus Clustering Reveals Distinct Subtypes and Survival Patterns. *Genes & Cancer*. 2010;1:152–163.

5. Brase JC, Schmidt M, Fischbach T, Sultmann H, Bojar H, Koelbl H, Hellwig B, Rahnenfuhrer J, Hengstler JG, Gehrmann MC. ERBB2 and TOP2A in breast cancer: a comprehensive analysis of gene amplification, RNA levels, and protein expression and their influence on prognosis and prediction. *Clin Cancer Res*. 2010;16:2391–2401.

6. Colman H, Zhang L, Sulman EP, et al. A multigene predictor of outcome in glioblastoma. *Neuro Oncol*. 2010;12:49–57.

7. Diaz-Moralli S, Tarrado-Castellarnau M, Alenda C, Castells A, Cascante M. Transketolase-like 1 expression is modulated during colorectal cancer progression and metastasis formation. *PLoS One*. 2011;6:e25323.

8. Donson AM, Birks DK, Barton VN, Wei Q, Kleinschmidt-Demasters BK, Handler MH, Waziri AE, Wang M, Foreman NK. Immune gene and cell enrichment is associated with a good prognosis in ependymoma. *J Immunol*. 2009;183:7428–7440.

9. Ebert C, von Haken M, Meyer-Puttlitz B, Wiestler OD, Reifenberger G, Pietsch T, von Deimling A. Molecular genetic analysis of ependymal tumors. NF2 mutations and chromosome 22q loss occur preferentially in intramedullary spinal ependymomas. *Am J Pathol*. 1999;155:627–632.

10. Freije WA, Castro-Vargas FE, Fang Z, Horvath S, Cloughesy T, Liau LM, Mischel PS, Nelson SF. Gene expression profiling of gliomas strongly predicts survival. *Cancer Res*. 2004;64:6503–6510.

11. Gilbertson RJ, Bentley L, Hernan R, Junttila TT, Frank AJ, Haapasalo H, Connelly M, Wetmore C, Curran T, Elenius K, Ellison DW. ERBB receptor signaling promotes ependymoma cell proliferation and represents a potential novel therapeutic target for this disease. *Clin Cancer Res*. 2002;8:3054–3064.

12. Guyotat J, Metellus P, Giorgi R, Barrie M, Jouvet A, Fevre-Montange M, Chinot O, Durand A, Figarella-Branger D. Infratentorial ependymomas: prognostic factors and outcome analysis in a multi-center retrospective series of 106 adult patients. *Acta Neurochir* (Wien) 2009;151:947–960.

13. Hasselblatt M. Ependymal tumors. *Recent Results Cancer Res*. 2009;171:51–66.

14. Healey EA, Barnes PD, Kupsky WJ, Scott RM, Sallan SE, Black PM, Tarbell NJ. The prognostic significance of postoperative residual tumor in ependymoma. *Neurosurgery*. 1991;28:666–671.

15. Jeuken JW, Sprenger SH, Gilhuis J, Teepen HL, Grotenhuis AJ, Wesseling P. Correlation between localization, age, and chromosomal imbalances in ependymal tumours as detected by CGH. *J Pathol*. 2002;197:238–244.

16. Johnson RA, Wright KD, Poppleton H, et al. Cross-species genomics matches driver mutations and cell compartments to model ependymoma. *Nature*. 2010;466:632–636.

17. Korshunov A, Golanov A, Timirgaz V. p14ARF protein (FL-132) immunoreactivity in intracranial ependymomas and its prognostic significance: an analysis of 103 cases. *Acta Neuropathol*. 2001;102:271–277.

18. Korshunov A, Golanov A, Timirgaz V. Immunohistochemical markers for prognosis of ependymal neoplasms. *J Neurooncol*. 2002;58:255–270.

19. Korshunov A, Neben K, Wrobel G, Tews B, Benner A, Hahn M, Golanov A, Lichter P. Gene expression patterns in ependymomas correlate with tumor location, grade, and patient age. *The American journal of pathology*. 2003;163:1721–1727.

20. Korshunov A, Witt H, Hielscher T, et al. Molecular staging of intracranial ependymoma in children and adults. *J Clin Oncol*. 2010;28:3182–3190.

21. Kuncova K, Janda A, Kasal P, Zamecnik J. Immunohistochemical prognostic markers in intracranial ependymomas: systematic review and meta-analysis. *Pathol Oncol Res*. 2009;15:605–614.

22. Louis DN, Ohgaki H, Wiestler OD, Cavenee WK, Burger PC, Jouvet A, Scheithauer BW, Kleihues P. The 2007 WHO classification of tumours of the central nervous system. *Acta Neuropathol*. 2007;114:97–109.

23. Massimino M, Buttarelli FR, Antonelli M, Gandola L, Modena P, Giangaspero F. Intracranial ependymoma: factors affecting outcome. *Future Oncol*. 2009;5:207–216.

24. Modena P, Lualdi E, Facchinetti F, et al. Identification of tumor-specific molecular signatures in intracranial ependymoma and association with clinical characteristics. *J Clin Oncol*. 2006;24:5223–5233.

25. Palm T, Figarella-Branger D, Chapon F, Lacroix C, Gray F, Scaravilli F, Ellison DW, Salmon I, Vikkula M, Godfraind C. Expression profiling of ependymomas unravels localization and tumor grade-specific tumorigenesis. *Cancer*. 2009;115:3955–3968.

26. Paulino AC. Radiotherapeutic management of intracranial ependymoma. *Pediatr Hematol Oncol*. 2002;19:295–308.

27. Peyre M, Commo F, Dantas-Barbosa C, et al. Portrait of ependymoma recurrence in children: biomarkers of tumor progression identified by dual-color microarray-based gene expression analysis. *PLoS One*. 2010;5:e12932.

28. Pezzolo A, Capra V, Raso A, Morandi F, Parodi F, Gambini C, Nozza P, Giangaspero F, Cama A, Pistoia V, Garre ML. Identification of novel chromosomal abnormalities and prognostic cytogenetics markers in intracranial pediatric ependymoma. *Cancer Lett*. 2008;261:235–243.

29. Phillips HS, Kharbanda S, Chen R, et al. Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis. *Cancer Cell*. 2006;9:157–173.

30. Ridley L, Rahman R, Brundler MA, et al. Multifactorial analysis of predictors of outcome in pediatric intracranial ependymoma. *Neuro Oncol*. 2008;10:675–689.

31. Robertson PL, Zeltzer PM, Boyett JM, et al. Survival and prognostic factors following radiation therapy and chemotherapy for ependymomas in children: a report of the Children's Cancer Group. *J Neurosurg*. 1998;88:695–703.

32. Rousseau E, Ruchoux MM, Scaravilli F, Chapon F, Vinchon M, De Smet C, Godfraind C, Vikkula M. CDKN2A, CDKN2B and p14ARF are frequently and differentially methylated in ependymal tumours. *Neuropathol Appl Neurobiol*. 2003;29:574–583.

33. Sparano JA, Paik S. Development of the 21-gene assay and its application in clinical practice and clinical trials. *J Clin Oncol*. 2008;26:721–728.

34. Suarez-Merino B, Hubank M, Revesz T, Harkness W, Hayward R, Thompson D, Darling JL, Thomas DG, Warr TJ. Microarray analysis of pediatric ependymoma identifies a cluster of 112 candidate genes including four transcripts at 22q12.1-q13.3. *Neuro Oncol*. 2005;7:20–31.

35. Taylor MD, Poppleton H, Fuller C, et al. Radial glia cells are candidate stem cells of ependymoma. *Cancer Cell*. 2005;8:323–335.

36. Verhaak RG, Hoadley KA, Purdom E, 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.

37. Witt H, Mack SC, Ryzhova M, et al. Delineation of two clinically and molecularly distinct subgroups of posterior fossa ependymoma. *Cancer Cell*. 2011;20:143–157.

38. Wolfsberger S, Fischer I, Hoftberger R, Birner P, Slavc I, Dieckmann K, Czech T, Budka H, Hainfellner J. Ki-67 immunolabeling index is an accurate predictor of outcome in patients with intracranial ependymoma. *Am J Surg Pathol*. 2004;28:914–920.

39. Zamecnik J, Snuderl M, Eckschlager T, Chanova M, Hladikova M, Tichy M, Kodet R. Pediatric intracranial ependymomas: prognostic relevance of histological, immunohistochemical, and flow cytometric factors. *Mod Pathol*. 2003;16:980–991.