

Outcome-based profiling of astrocytic tumours identifies prognostic gene expression signatures which link molecular and morphology-based pathology

CHRISTIAN BEETZ¹, SVEN BERGNER¹, STEFAN BRODOEHL¹, MICHAEL BRODHUN², CHRISTIAN EWALD³, ROLF KALFF³, JUTTA KRÜGER⁴, STEPHAN PATT², MICHAEL KIEHNTOPF¹ and THOMAS DEUFEL¹

¹Institut für Klinische Chemie und Laboratoriumsdiagnostik, ²Institut für Pathologie, ³Klinik für Neurochirurgie, Universitätsklinikum, Jena; ⁴Klinik für Neurochirurgie, HELIOS Klinikum, Erfurt, Germany

Received May 18, 2006; Accepted July 12, 2006

Abstract. Astrocytomas are intracranial malignancies for which invasive growth and high motility of tumour cells preclude total resection; the tumours usually recur in a more aggressive and, eventually, lethal form. Clinical outcome is highly variable and the accuracy of morphology-based prognostic statements is limited. In order to identify novel molecular markers for prognosis we obtained expression profiles of: i) tumours associated with particularly long recurrence-free intervals, ii) tumours which led to rapid patient death, and iii) tumour-free control brain. Unsupervised data analysis completely separated the three sample entities indicating a strong impact of the selection criteria on general gene expression. Consequently, significant numbers of specifically expressed genes could be identified for each entity. An extended set of tumours was then investigated by RT-PCR targeting 12 selected genes. Data from these experiments were summarised into a sample-specific index which assigns tumours to high- and low-risk groups as successfully as does morphology-based grading. Moreover, this index directly correlates with definite survival suggesting that integrated gene expression data allow individualised prognostic statements. We also analysed localisation of selected marker transcripts by *in situ* hybridization. Our finding of cell-specificity for some of these outcome-determining genes relates global expression data to the presence of morphological correlates of tumour behaviour and, thus, provides a link between morphology-based and molecular pathology. Our identification of expression signatures that are associated individually with clinical outcome confirms the prognostic

relevance of gene expression data and, thus, represents a step towards eventually implementing molecular diagnosis into clinical practice in neuro-oncology.

Introduction

Malignant glioma is the most common primary brain tumour in adults. Classification is built on resemblance of tumour cells to certain types of glial cells; generally, ependymomas, oligodendrogliomas and astrocytomas are recognised (1). Complete resection of these highly infiltrating malignancies is usually not achieved and recurrent relapses are eventually fatal (2). Nevertheless, outcome is highly variable within each class and several morphology-based grading systems have been developed to account for this (1,3,4). The prognostic validity of grading, however, is still limited due to considerable morphological and clinical variability, partial overlap between grades, and large inter-observer discordance in the scoring of the relevant parameters (5-7).

The potential of molecular criteria to refine or even replace existing glioma classification and/or grading schemes has repeatedly been discussed (6,8). A prominent example, the determination of loss vs. retention of chromosome 1p to predict outcome in oligodendrogliomas (9), is about to enter clinical practice. Amongst the novel technologies available to screen for additional molecular characteristics of pre-defined tumour entities, microarray-based transcriptome analysis is the most widely applied. A number of pertinent studies on gliomas have yielded promising results: it was shown that statistically significant differences exist between the gene expression profiles of: i) normal brain and gliomas (10-12), ii) different glioma entities (13-15), iii) different grades within one entity (13,14,16-23), iv) histologically identical gliomas with different genetic background (24-26), v) tumours with differing contents of necrosis (27), and vi) primary and recurrent tumours (28). Whereas these studies were primarily focused on the biological implications of their findings, the prognostic relevance of the groups discriminated was in some cases shown by secondary analyses (15,26). More recently, expression profiles have also primarily been correlated to prognosis. This has led to the identification of prognostic gene expression markers for high-grade gliomas (29,30) and glioblastomas (31-33).

Correspondence to: Professor Thomas Deufel, Institut für Klinische Chemie und Laboratoriumsdiagnostik, Universitätsklinikum Jena, D-07740 Jena, Germany
E-mail: thomas.deufel@med.uni-jena.de

Key words: astrocytoma, expression profiling, outcome, prognosis

Table I. Clinico-pathological data for samples analysed in this study.

Survival in months	Survival status at last contact	Method of analysis P: by RT-PCR, A: on array (entity)	Sex	Age at diagnosis	WHO grade
3	Dead	P/A(O2)	Male	48	IV
4	Dead	P/A(O2)	Female	77	IV
4	Dead	P	Male	77	IV
5	Dead	P	Male	46	IV
8	Dead	P	Male	52	IV
9	Dead	P/A(O2)	Male	73	IV
11	Dead	P	Male	73	IV
11	Dead	P/A(O2)	Female	64	IV
12	Dead	P/A(O2)	Female	62	IV
13	Dead	P	Male	63	IV
14	Alive	P	Female	22	II
14	Dead	P	Male	47	III
16	Alive	P	Male	33	II
16	Dead	P	Male	40	IV
17	Dead	P	Female	45	IV
21	Alive	P	Male	13	III
23	Alive	P	Female	52	IV
28	Alive	P/A(O1)	Male	17	II
29	Dead	P	Male	41	IV
30	Alive	P	Male	51	II
32	Alive	P/A(O1)	Female	32	II
33	Alive	P/A(O1)	Female	34	II
36	Alive	P	Male	36	III
40	Alive	P/A(O1)	Male	31	II
46	Alive	P	Female	25	II

In the present study, microarray-based expression profiling was applied to investigate control brain and malignant astrocytomas of all grades with the primary objective of discerning tumour entities defined by extreme, i.e. favourable vs. poor, outcomes. The aim of these experiments was to identify specific molecular signatures, the prognostic relevance of which was then to be tested by an independent methodology on a larger set of samples. Since the clinical correlate of interest, i.e. survival after initial diagnosis, is a continuous measure (34), we further hypothesised that a continuous numerical index developed from the expression data is more appropriate than the common assignment of class labels. Finally, we used *in situ* hybridization in order to link marker gene expression to recognised morphological correlates of astrocytoma prognosis.

Materials and methods

Sampling and patient selection. This study has been approved by the Ethics Committee at the medical faculty of the Friedrich-Schiller-Universität Jena, Germany. Tissue samples were collected from patients undergoing brain tumour surgery by neurosurgery departments at Universitätsklinikum Jena and HELIOS-Klinikum Erfurt. Part of the resected tissue was

used for neuropathology routine diagnosis according to the WHO classification criteria (1); the remaining portion was immediately snap-frozen in liquid nitrogen and stored at -80°C . Clinico-pathological data are summarised in Table I. Cases conceptually suitable for microarray analysis were selected by browsing the post-operative clinical records of patients diagnosed with malignant astrocytoma who suffered tumour-related death within 12 months following initial diagnosis (poor outcome; henceforth referred to as sample entity O2) and for patients without aggressive recurrence for more than 2 years (favourable outcome, O1); control brain samples (O0) were obtained by screening glioma resections for non-tumourous areas.

Sample selection and extraction of RNA. Frozen tissue was chopped into pieces of varying size and shape. Using a cryostat for successively removing material from opposing sides, these pieces were subsequently trimmed such as to represent flat discs of $\sim 30\text{-}50$ mg weight and ~ 2 mm height. The ultimate $8\text{-}\mu\text{m}$ sections from each side, regarded representative for the corresponding disc, were subjected to standard haematoxylin and eosin staining (Fig. 1A); cellular composition was evaluated by an experienced neuropathologist (Stephan Patt).

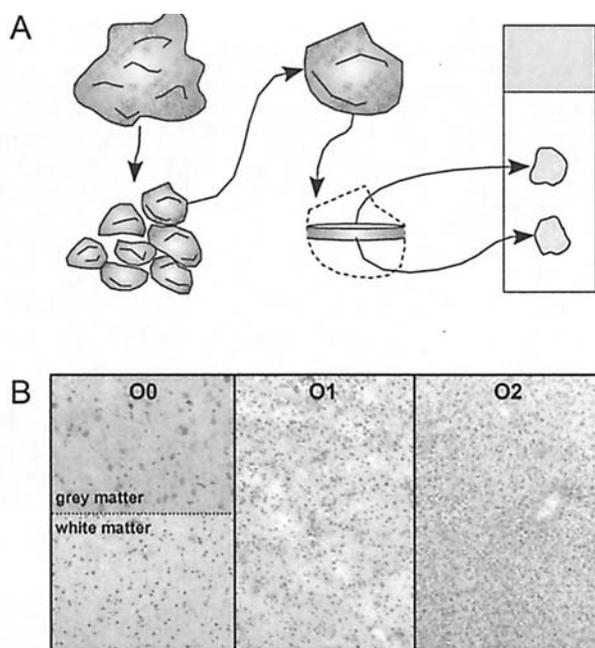


Figure 1. Sample characterisation. (A) Snap-frozen resected tissue was chopped into smaller pieces which served to prepare flat discs; cellular composition of discs was assessed on H&E-stained upper- and lower-side sections. (B) Representative sections for O0, O1, and O2 samples showing homogeneous cellular composition. Note that O0 samples were chosen to contain both grey and white matter.

Ideally, tumour samples would be composed of >90% evenly distributed, non-necrotic neoplastic cells, whereas control brain would contain equal proportions of cortex and white matter lacking any tumour-related morphology. Only discs for which both sections fulfilled the above criteria (Fig. 1B) were employed in further analyses. Frozen tissue was homogenised and total RNA extracted by use of a commercial kit (RNeasy® Mini kit; Qiagen, Hilden, Germany) following the manufacturer's instructions. RNA integrity and concentration was examined on an Agilent 2100 Bioanalyzer (Agilent Technologies; Palo Alto, CA, USA) using the RNA 6.000 LabChip kit (Agilent Technologies) according to the manufacturer's instructions.

Oligonucleotide array analysis. Probe labelling, hybridization to the U95Av2 array (Affymetrix), and scanning were conducted at the microarray core facility of the Interdisziplinäres Zentrum für Klinische Forschung (IZKF) Leipzig (Faculty of Medicine, University of Leipzig) as described previously (35). Scanned images were processed using software Microarray Suite 5.0 (MAS5.0, Affymetrix): absolute analyses were carried out for each of the 12 arrays to derive gene-specific detection calls (statistics-based assignment of absent, marginal or present) and signals (numerical expression value); linear scaling was implemented in order to normalise data for inter-array comparisons. We also calculated the gene-specific change p-values (CPVs) from all pair-wise array-to-array comparisons as implemented in MAS5.0. Note that this value is different from, and partly independent of, the ANOVA-derived p-value introduced below.

Analysis of array data and generation of candidate gene lists. Data points for spiked-in transcripts, for probe sets used

for quality control, and for genes scored absent in all 12 experiments were removed; signal values of the 8499 remaining genes were imported into the Cluster and TreeView software package (36). First, data were log₂-transformed and median centred; for subsequent analyses, either the complete set of 8499 genes or a subset of 656 selected genes (see below for selection criteria) was used. Arrays were hierarchically clustered using the complete linkage of Euclidean distance measures option. One-way ANOVA models were fit to the data in order to identify genes for which expression was specific (at $p < 0.0005$) to entities O0, O1, and O2; overexpressed genes were separated from underexpressed genes by applying k-means clustering ($k=2$, maximum iterations=100). This resulted in six candidate gene lists coherently termed O0_low, O0_high, O1_low, O1_high, O2_low, and O2_high. In addition, entries common to the O0-specific and the O2-specific lists (at $p=0.005$) were compiled to yield two additional lists which contain genes with continuously decreasing and increasing expression along the O0-O1-O2 spectrum.

Semiquantitative RT-PCR. Genes that were comprised in the above defined gene lists and consistently showed CPVs close to one or to zero (depending on the direction of change) were selected for RT-PCR. Primers were designed according to the criteria that they i) preferentially target 3' mRNA regions lacking orthologous sequences but being present in all known isoforms, ii) span intronic sequences, and iii) amplify ~300-600 bp; all oligonucleotides were obtained from MWG-Biotech (Ebersberg, Germany). After optimization for specificity and sub-cloning of the PCR-product followed by confirmatory sequencing, a total of 12 genes (for identity see Fig. 4B) were chosen for further analysis on an expanded set of samples. cDNA was prepared from 500 ng total RNA using an iScript™ cDNA synthesis kit (Bio-Rad; Hercules, CA, USA) according to the manufacturer's instructions. Templates for PCR were prepared as 1:15 dilutions of the cDNA reactions. *ACTB* (β -actin, NM_001101) was used as the reference house-keeping gene; in a pilot experiment, 22 PCR cycles were determined as to represent the exponential phase of amplification for every template when targeting *ACTB*. For the other genes, a correspondingly optimised number of PCR cycles was established from PCRs on pooled cDNAs. The intensity of bands from single reactions was measured densitometrically, normalised against the intensity of the *ACTB* band, and scaled to a mean of 1 across all samples to derive gene- and sample-specific expression values.

Statistical analysis of RT-PCR data. Expression values for samples analysed on the arrays were used to show the degree of correspondence between array and RT-PCR-derived expression data. The difference in expression values between entities (according to the underlying candidate list) was tested for significance ($p < 0.05$) using the two-sided exact Mann-Whitney U test as implemented in the SSPS software package (Scientific Software for Professional Statistics, Chicago, IL, USA).

In order to relate RT-PCR data to survival, we intended to subsume the expression values for all 12 genes into a single, sample-specific measure. To this end, we first ranked expression values for each gene; positive or negative correlation with outcome, suggested by the array data as increase or

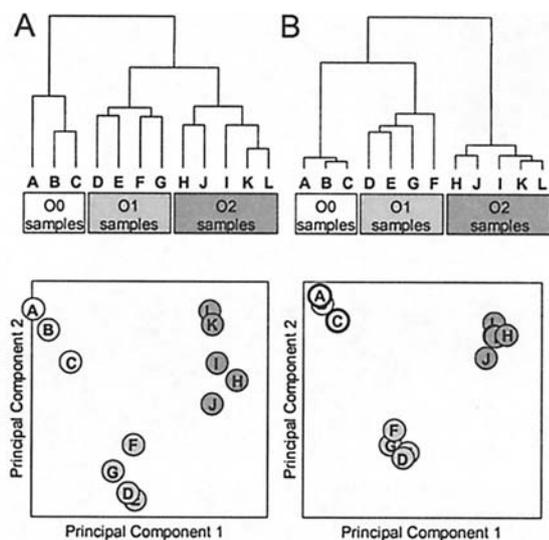


Figure 2. Hierarchical clustering (HC, top) and principal component analysis (PCA, bottom) of expression profiles. (A) Unsupervised analysis using all 8499 genes recovers the pre-defined sample entities as unique clusters in HC, with O1 samples weakly associating to O2 samples; entities are also well-separated by PCA. (B) Supervised analysis, i.e. the use of only 656 selected genes (see text for selection criteria), generally strengthens this structure. In HC, however, the position of individual O1 samples is changed and the whole entity is now weakly associated with O0. Note that also in PCA a clear affiliation of O1 samples to either O0 or O2 is lacking and that principal component 1 (explaining most variability in the data) indicates an intermediate position for O1.

decrease from O1 to O2, was implemented such that highest ranks correspond to O1-like expression. These gene-specific ranks were then added up for every sample (rank-sum) yielding

an integer value to which each of the 12 genes contributes equally (henceforth referred to as integrated expression index, IEI). The prognostic relevance of IEI was compared to that of morphology-based grading by generating Kaplan-Meier plots (37): partition by grading was implemented by combining grade II ($n=8$) and grade III ($n=3$) tumours and comparing them to grade IV tumours ($n=14$); the cut-off for IEI was set at 160 to yield equally sized low- and high-risk groups, i.e. $n=11$ vs. $n=14$, respectively. In addition, survival was plotted against IEI for all tumours with defined end-point data.

In situ hybridization. Gene-specific sequences of ~400 bp were cloned into pCR[®]II-TOPO[®] (Invitrogen, Karlsruhe, Germany) in both orientations to enable the use of a single promoter (*SP6*) for generation of sense and anti-sense probes. Digoxigenin-labelled cRNA was prepared using DIG RNA labelling kit (Roche Diagnostics GmbH, Mannheim, Germany) according to the manufacturer's recommendations. Cryosections of 12 μ m were probed essentially as described previously (38) with hybridization at 65°C and washing at 60°C as the sole modifications. Visualisation utilised a DIG nucleic acid detection kit (Roche Diagnostics GmbH) as recommended by the manufacturer.

Results

Stringent selection criteria identify specimens with well-defined cellular composition. Based on post-operative records of patients diagnosed with astrocytoma, 14 cases fulfilling the aforementioned clinical criteria for being considered for microarray analysis were selected. Careful inspection of corresponding tissue (Fig. 1A) led us to exclude a considerable

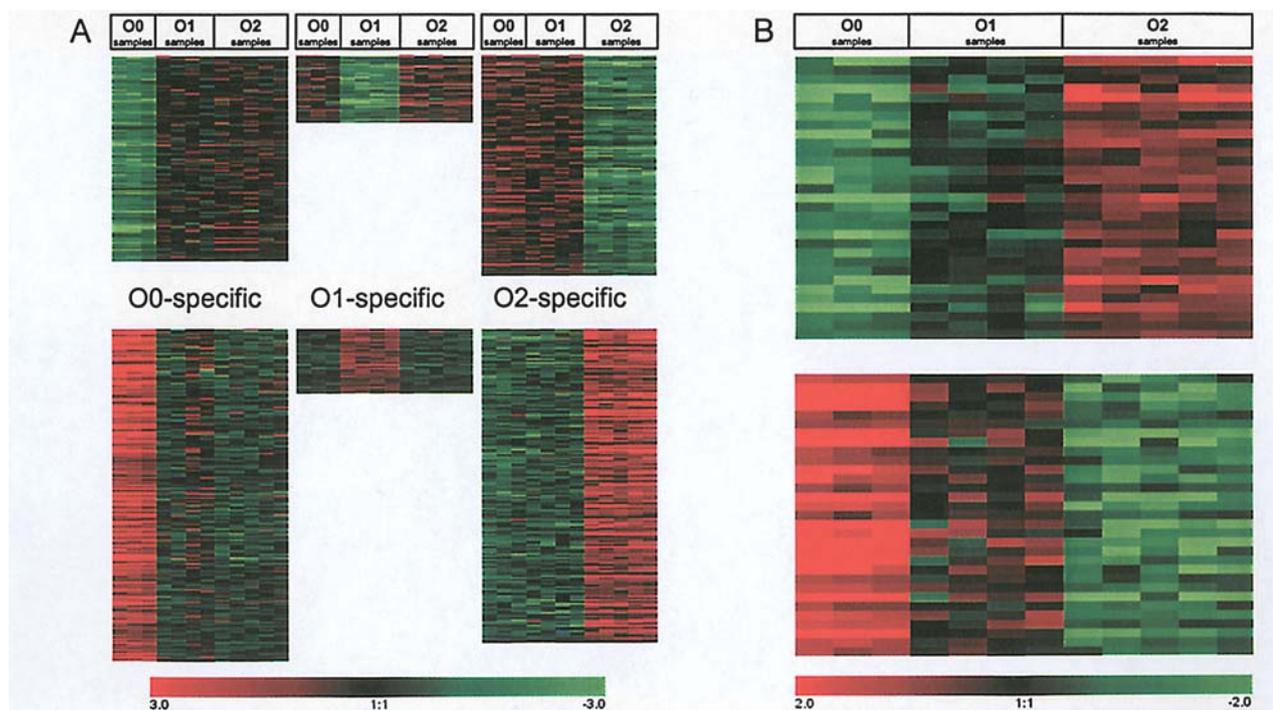


Figure 3. Gene expression associated with sample entities defined by clinical outcome. Columns represent samples, rows represent genes. Relative expression is colour-coded continuously from low (green) to high (red). (A) Genes significantly ($p=0.0005$, One-way ANOVA) overexpressed (top) or underexpressed (bottom) in sample entities O0, O1 and O2. (B) Genes for which expression increases (top) or decreases (bottom) along the O0-O1-O2 gradient (lists generated as overlap of O0-specific and O2-specific genes at $p=0.005$, One-way ANOVA).

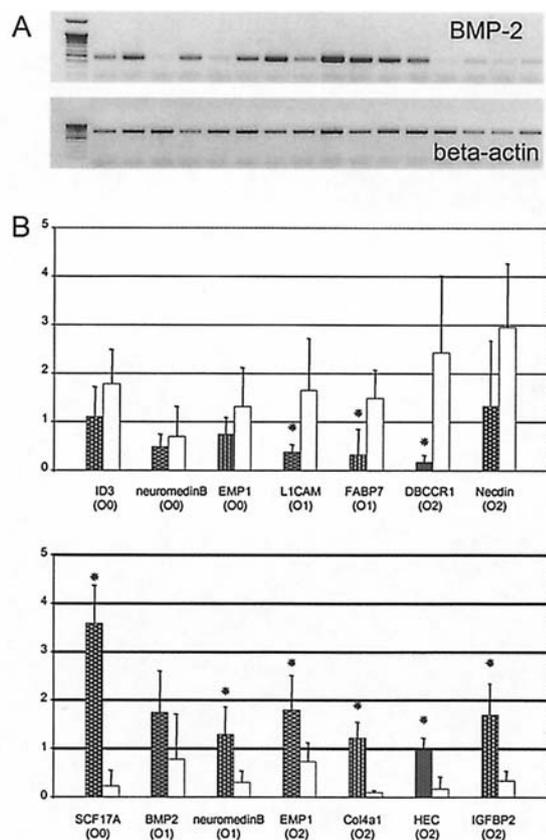


Figure 4. RT-PCR results and relation to array data. (A) Representative RT-PCR results. Sample- and gene-specific relative expression values were derived from raw images by densitometric analysis of band intensities followed by normalisation to β -actin and scaling to a mean of 1 for every gene across all samples. (B) Data for samples analysed on the arrays and re-analysed by RT-PCR. Mean relative expression values for genes underrepresented or overrepresented (upper and lower chart, respectively) by expression profiling in the indicated entities. Asterisks denote statistical significance ($p < 0.05$; two-sided exact Mann-Whitney U test).

number of tumours: samples of homogeneous cellular composition (Fig. 1B) and sufficient size could be identified for four of seven O1 cases and five of seven O2 cases; ~20 glioma resections had to be screened to identify three O0 samples. The cases eventually employed for microarray analysis are indicated in Table I which presents the clinico-pathological data of all samples used in this study.

Expression profiles discriminate between sample entities defined by outcome. In order to uncover a structure inherent to the complete set of expression data we first applied unsupervised analyses, i.e. no information as to the identity of the samples was incorporated. Hierarchical clustering using the complete set of 8499 genes resulted in the separation of O0 samples from O1 and O2 samples, i.e. control brain tissue from tumours; the two sub-clusters within the tumour branch exactly coincided with O1 and O2 (Fig. 2A, top). Principal component analysis with the same set of genes revealed the data to be highly structured; the single (two, three) most informative component(s) explain 31% (48%, 58%) of the variation in the data compared to the 8.3% (16.7%, 25%) expected from a structure-less dataset of equal size. A two-dimensional plot of the first two principal components

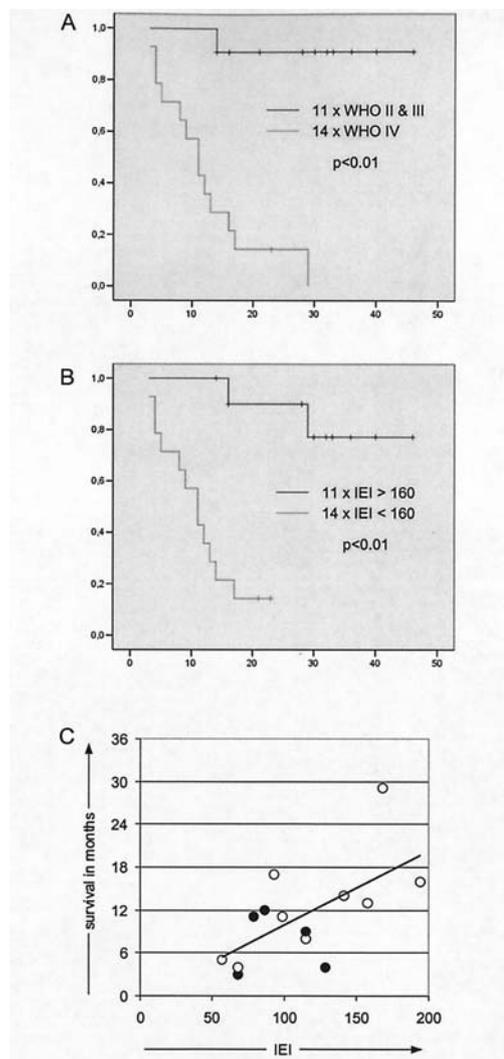


Figure 5. Prognostic value of morphological and gene expression data. (A) Kaplan-Meier plot based on WHO grading. (B) Kaplan-Meier plot derived from grouping tumours according to an RT-PCR-based integrated expression index (IEI; see Materials and methods for details). (C) Correlation between IEI and definite survival associated with individual tumours. Filled circles represent the cases that had been analysed on the array as entity O2.

(PC1 and PC2) visualises separation of O0, O1, and O2, with O1 samples in an intermediate position according to PC1 but also possessing specificity according to PC2 (Fig. 2A, bottom).

Sample entities defined by outcome are characterised by specific sets of differentially expressed genes. Having shown that the overall structure of the microarray data supports distinction between O0, O1 and O2, we aimed at identifying the genes underlying such structure. To this end, One-way ANOVA models were fit to pair-wise comparisons of respective groups of samples. Fig. 3A visualises the expression of 656 genes identified at $p < 0.0005$. The number of genes specific for each entity is well above the value expected by chance alone (~4 per entity) but highly varies ($n=72$ for O1 compared to $n=293$ and $n=291$ for O0 and O2, respectively). Gene lists of significant lengths, with those for O1 being shortest, were also obtained by varying the cut-off from $p < 0.00025$ to $p < 0.01$ (data not shown).

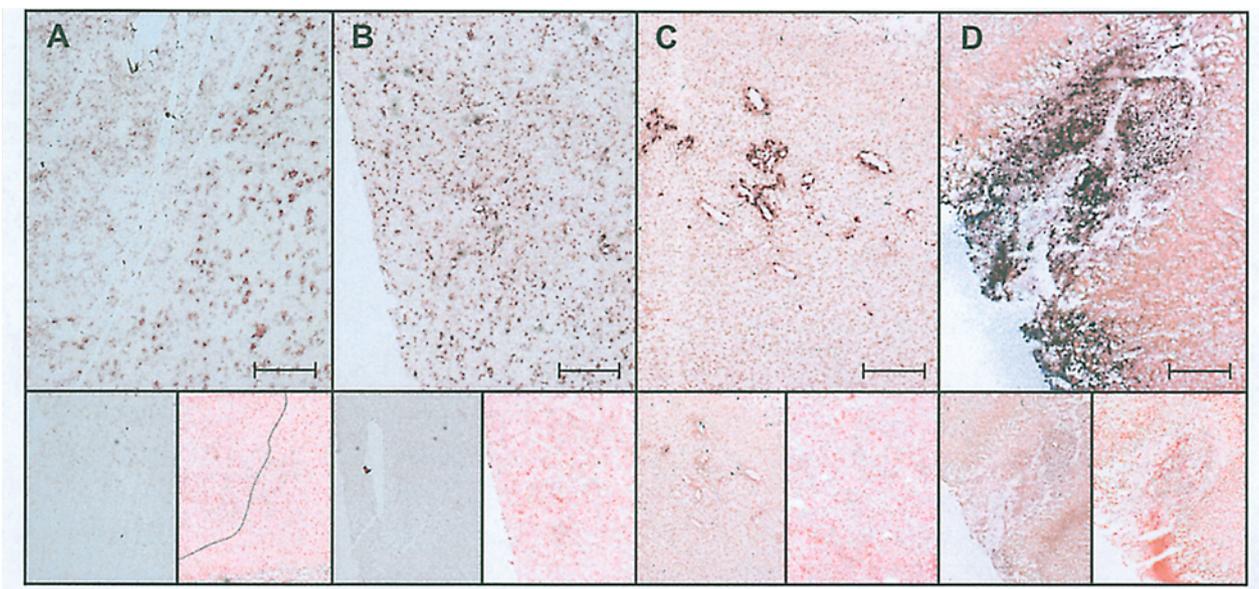


Figure 6. Cellular localisation of selected transcripts. Images taken after *in situ* hybridization with anti-sense cRNA (large frames) and sense cRNA (lower left frames) and after H&E staining (lower right frames). (A) Expression of *SLC17A* (contained in O0_high) is found in large cell bodies of grey matter presumably representing neurons (border grey/white matter stippled in H/E image). (B) *HAPIP* (O0_high) is expressed in smaller, most likely glial cells in control white matter. (C) *COL4A1* (O2_high) expression is restricted to endothelial cells in the vascularised areas of a glioblastoma sample. (D) Strong *EMP1* (O0_low and O2_high) expression is detected in glioblastoma cells embedded in a necrotic area. Scale bars, 200 μ m.

We next used the 656 genes from the list generated at $p < 0.0005$ to repeat the above similarity analyses in a supervised manner. Again, the three sample entities O0, O1 and O2 positioned to unique branches in hierarchical clustering (Fig. 2B, top). While topology was conserved and uniformity greatly enhanced within O0 and O2, this did not apply to O1 (compare Fig. 2A, top and Fig. 2B, top). Moreover, the very position of the O1 branch was changed. However, the new affiliation of O1 to O0 was weak, just like its affiliation to O2 in the unsupervised analysis. A two-dimensional plot of the first two principal components derived from PCA on the 656 selected (Fig. 2B, bottom) genes revealed a similar picture, i.e. greatly enhanced separation of O0, O1 and O2 but no clear affinity of O1 to either O0 or O2.

We next wanted to know whether the intermediate position of the O1 samples suggested in all previous analyses would also be reflected in the expression pattern of specific genes. Matching the lists with O0- and O2-specific genes (at $p < 0.005$) identified 62 genes with steadily increasing or decreasing expression when assuming an O0-O1-O2 continuity (Fig. 3B) which is far more than expected by chance (~4 genes); this number dropped to 44 and 4 when setting O0 or O2 as intermediate, respectively.

RT-PCR data confirm array analyses and further support a correlation of gene expression with survival. In order to validate the array results we choose 12 genes from the above generated lists and performed semi-quantitative RT-PCR (Fig. 4A) on newly prepared RNA from the tumours used for array analysis. Statistically significant differences regarding expression values were revealed for 9 of the 14 comparisons; in those not reaching significance, the tendency, i.e. relation of the means, was still in accordance with the array data (Fig. 4B).

Knowing that our RT-PCR data, by largely overlapping with the array results, are also associated with outcome in

these samples, we extended survival analyses to include additional cases. The RT-PCR-based IEI as depicted in Materials and methods separates long-term from short-term survivors as successfully as does morphology-based, classical grading (Fig. 5A and B). Moreover, plotting survival against IEI for cases with defined end-point data reveals a positive correlation that was most evident for the non-array samples which include long-surviving cases.

Distinctive expression signatures derived by global analysis are accounted for by the presence/absence of specific cell populations. Both control brain tissue and tumour samples are heterogeneous in their cellular composition and we were interested to see to what extent the global gene expression signature can be attributed to the presence of specific cell types or morphological characteristics present in the tissue sample. To this end we chose to investigate localisation of transcripts specific for either O0 or O2 samples. Two transcripts enriched in O0 are expressed in different cell types in control brain: the *SLC17A* probe marks large cells in the cortex which presumably represent neurons (Fig. 6A); *HAPIP* is specific to small, most likely glial cells (Fig. 6B). A similar association to specific cell types is seen for two genes overrepresented in O2: transcript for *Col4A1* is exclusively found in the endothelial cells lining tumour-specific blood vessels (Fig. 6C), whereas *EMP1* is strongly expressed in tumour cells neighbouring necrotic areas in glioblastoma samples (Fig. 6D).

Discussion

Our study was aimed at identifying gene expression signatures which are associated with prognosis in astrocytic tumours and, hence, possess potential diagnostic value. To this end we compared the expression profiles of control brain tissue and two tumour sample entities defined by extreme, namely

favourable and poor, outcome. With this definition of groups solely on the basis of post-operative course, our approach conceptually differs from that taken in many earlier studies (21,29) as it builds on clinical rather than histological information. In general, reports on good correlation between histological parameters and a variety of molecular features of astrocytoma (7,13,39,40) can help in understanding the biology of these tumours; yet, regarding clinical prognosis the real challenge would be to identify new criteria which more accurately predict clinical outcome rather than to correctly classify tumours into already existing subgroups (34,41).

Along this line, a number of more recent studies on high-grade gliomas have directly related expression profiles to survival; very impressively, they succeeded in defining relevant markers even within these entities generally associated with poor prognosis (30,31,33). The fact that some of the same candidates (e.g. *FABP7*, *BMP2*) also come up in our study corroborates these earlier results and further validates our approach. Furthermore, by being based on the inclusion of low-grade tumours our data suggest an even broader significance for some of the previously reported markers. Additional examples of genes known to be differentially regulated during astrocytoma progression and identified in this study include *VEGF* (42; contained in O2_high), *IGFBP2* (16; O2_high), *tenascin-C* (43; O0_low), *SPARC* (44; O0_low) and *vimentin* (17; O0_low and O2_high).

Meticulous sample characterisation is a critical step in large-scale molecular screens on a limited number of cases; it ascertains intra-sample homogeneity as well as uniformity of samples within entities of interest. Yet, assessment of the former point in related studies is either not reported (10,11,26) or only dealt with by reviewing a single H&E stain (14,19,29). This is surprising when considering the infiltrating growth and, as a consequence, the compositional diversity of resections as typical of malignant gliomas. We optimised the evaluation of cellular composition by shaping samples such as to have a high ratio of analysable sample surface to non-analysable sample core (Fig. 1A). The high drop-out rate resulting from this evaluation in our protocol just underscores the necessity of including such a step. Uniformity within pre-defined sample entities was achieved by non-overlap of the defining criteria; this is similar to pre-selecting only samples with well-agreed, 'typical' histology when comparing diagnostic groups (15,29) or when analysing diagnostically relevant areas within a given tissue (32).

Ideally, microarray experiments are initiated by one of two objectives which, in turn, determine subsequent analytical procedure: class discovery is usually carried out under unsupervised conditions whereas supervised methods are employed in class comparison/prediction (34). Despite our approach falling among the latter, we initially addressed the structure immanent to the complete set of data by applying hierarchical clustering and principal component analysis in an unsupervised manner. To our surprise, the clusters generated completely correspond to the pre-defined sample entities. This may in part be interpreted a random effect as we note that complete congruence between unsupervised grouping and a given pre-definition of sample entities is usually reported in studies on relatively small numbers of samples (19,23; this

study). Alternatively, of all factors influencing the transcriptome it may in fact be prognosis leaving the strongest footprints. In support of this notion, glioblastomas from long-surviving patients were placed close to lower-grade samples by unsupervised analysis (22,39) and oligodendrogliomas with a good prognosis were more similar to normal brain than to histologically similar tumours with a bad prognosis (15,25). Finally, and not mutually exclusive, the strict tumour and sample selection criteria discussed above may indeed have succeeded in creating more uniform entities.

The samples used in the microarray experiments comprise three distinct entities along a gradient of clinically defined disease severity which ranges from healthy (O0) to highly unfavourable outcome (O2). An intermediate position of O1 is also reflected in several aspects of the corresponding expression profiles. Firstly, global data analysis unstably assigns O1 to either O0 or O2 while never grouping together O0 and O2 (Fig. 2). Secondly, the number of genes for which expression is entity-specific is considerably higher for O0 and O2 than for O1 (Fig. 3A). Finally, when O1 is set as intermediate (Fig. 3B), the number of genes found to steadily decrease or increase is higher than the number obtained when this position is assigned to O0 or O2. A similar 'in-between phenomenon' of expression profiles has been described previously when comparative expression profiling employing all classes of gliomas failed to identify specifically expressed genes for oligodendrogliomas with intact 1p and for low-grade astrocytomas (25); interestingly, amongst the groups analysed, these two entities are generally associated with intermediate prognosis again. Taken together, such data suggest clinical malignancy to be encoded in a continuous spectrum in expression profiles. We have tested this hypothesis by defining the IEI described above as a sample-specific continuous value and show that this value not only divides the samples into two discrete groups with significantly different outcome (Fig. 5B), but that it also correlates with survival in individual cases (Fig. 5C). Together with the successful application of a similar analytical strategy recently reported for glioblastomas (31), this argues for a potential of expression data to highly individualise prognosis in gliomas.

The specific approach of our study was to use expression profiling on a defined set of tumours to derive prognostic gene expression patterns. It was then of particular interest to us to learn which characteristics would help a given gene within this pattern to qualify as a marker. It became evident from our data that this was not only its being expressed in the tumour cells but that expression differences pertained to various other cell types present in the samples. When we characterised cellular localisation of selected transcripts we found glia-specific as well as neuron-specific genes amongst those enriched in control brain, which is consistent with previously reported findings (33). Among the genes specific to the high malignancy samples, one was found to be strongly expressed in the vicinity of necrotic areas, the other marked blood vessel walls; with both necrosis and vascularisation being hallmarks of aggressive gliomas (2), these findings directly relate the expression of certain genes to well-known morphological features of malignancy. Incidentally, the discriminative power of such markers might be lost if one tried, e.g. by microdissection procedures, to enrich for

tumour cells in the substrate for molecular analysis. On the contrary, we might expect that the complexity of information contained in tissue morphology that enables conventional neuropathology grading can indeed be reflected by data obtained by molecular analysis alone on a complex diagnostic sample.

Acknowledgments

We thank P. Süptitz and Dr K. Krohn from the microarray core facility of the Interdisziplinäres Zentrum für Klinische Forschung (IZKF) Leipzig for excellent service as regards the technical steps of the microarray experiments. This work was supported by the IZKF Jena, FS1 'Klinisch-experimentelle Onkologie', TP3.8.

References

- Kleihues P and Cavenee WK: Pathology and Genetics - Tumours of the Central Nervous System. IARC Press, Lyon, 2000.
- Collins VP: Gliomas. *Cancer Surv* 32: 37-51, 1998.
- Kernohan JW, Mabon RF, Svien HJ and Adson AW: A simplified classification of gliomas. *Proc Staff Meet Mayo Clin* 24: 71-75, 1949.
- Daumas-Duport C, Scheithauer B, O'Fallon J and Kelly P: Grading of astrocytomas. A simple and reproducible method. *Cancer* 62: 2152-2165, 1988.
- Coons SW, Johnson PC, Scheithauer BW, Yates AJ and Pearl DK: Improving diagnostic accuracy and interobserver concordance in the classification and grading of primary gliomas. *Cancer* 79: 1381-1393, 1997.
- Louis DN, Holland EC and Cairncross JG: Glioma classification: a molecular reappraisal. *Am J Pathol* 159: 779-786, 2001.
- Ueki K, Nishikawa R, Nakazato Y, *et al.*: Correlation of histology and molecular genetic analysis of 1p, 19q, 10q, tp53, egfr, cdk4, and cdkn2a in 91 astrocytic and oligodendroglial tumors. *Clin Cancer Res* 8: 196-201, 2002.
- Caskey LS, Fuller GN, Bruner JM, Yung WK, Sawaya RE, Holland EC and Zhang W: Toward a molecular classification of the gliomas: histopathology, molecular genetics, and gene expression profiling. *Histol Histopathol* 15: 971-981, 2000.
- Cairncross JG, Ueki K, Zlatescu MC, *et al.*: Specific genetic predictors of chemotherapeutic response and survival in patients with anaplastic oligodendrogliomas. *J Natl Cancer Inst* 90: 1473-1479, 1998.
- Huang H, Colella S, Kurrer M, Yonekawa Y, Kleihues P and Ohgaki H: Gene expression profiling of low-grade diffuse astrocytomas by cdna arrays. *Cancer Res* 60: 6868-6874, 2000.
- Markert JM, Fuller CM, Gillespie GY, *et al.*: Differential gene expression profiling in human brain tumors. *Physiol Genomics* 5: 21-33, 2001.
- Madden SL, Cook BP, Nacht M, *et al.*: Vascular gene expression in non-neoplastic and malignant brain. *Am J Pathol* 165: 601-608, 2004.
- Kim S, Dougherty ER, Shmulevich L, *et al.*: Identification of combination gene sets for glioma classification. *Mol Cancer Ther* 1: 1229-1236, 2002.
- Shai R, Shi T, Kremen TJ, *et al.*: Gene expression profiling identifies molecular subtypes of gliomas. *Oncogene* 22: 4918-4923, 2003.
- Huang H, Okamoto Y, Yokoo H, *et al.*: Gene expression profiling and subgroup identification of oligodendrogliomas. *Oncogene* 23: 6012-6022, 2004.
- Fuller GN, Rhee CH, Hess KR, *et al.*: Reactivation of insulin-like growth factor binding protein 2 expression in glioblastoma multiforme: A revelation by parallel gene expression profiling. *Cancer Res* 59: 4228-4232, 1999.
- Sallinen SL, Sallinen PK, Haapasalo HK, *et al.*: Identification of differentially expressed genes in human gliomas by DNA microarray and tissue chip techniques. *Cancer Res* 60: 6617-6622, 2000.
- Rickman DS, Bobek MP, Misek DE, *et al.*: Distinctive molecular profiles of high-grade and low-grade gliomas based on oligonucleotide microarray analysis. *Cancer Res* 61: 6885-6891, 2001.
- Watson MA, Perry A, Budhara V, Hicks C, Shannon WD and Rich KM: Gene expression profiling with oligonucleotide microarrays distinguishes world health organization grade of oligodendrogliomas. *Cancer Res* 61: 1825-1829, 2001.
- Hunter S, Young A, Olson J, *et al.*: Differential expression between pilocytic and anaplastic astrocytomas: identification of apolipoprotein d as a marker for low-grade, non-infiltrating primary CNS neoplasms. *J Neuropathol Exp Neurol* 61: 275-281, 2002.
- Tanwar MK, Gilbert MR and Holland EC: Gene expression microarray analysis reveals ykl-40 to be a potential serum marker for malignant character in human glioma. *Cancer Res* 62: 4364-4368, 2002.
- Godard S, Getz G, Delorenzi M, *et al.*: Classification of human astrocytic gliomas on the basis of gene expression: a correlated group of genes with angiogenic activity emerges as a strong predictor of subtypes. *Cancer Res* 63: 6613-6625, 2003.
- Khatua S, Peterson KM, Brown KM, *et al.*: Overexpression of the egfr/fkbp12/hif-2alpha pathway identified in childhood astrocytomas by angiogenesis gene profiling. *Cancer Res* 63: 1865-1870, 2003.
- Gutmann DH, Hedrick NM, Li J, Nagarajan R, Perry A and Watson MA: Comparative gene expression profile analysis of neurofibromatosis 1-associated and sporadic pilocytic astrocytomas. *Cancer Res* 62: 2085-2091, 2002.
- Mukasa A, Ueki K, Matsumoto S, *et al.*: Distinction in gene expression profiles of oligodendrogliomas with and without allelic loss of 1p. *Oncogene* 21: 3961-3968, 2002.
- Mischel PS, Shai R, Shi T, *et al.*: Identification of molecular subtypes of glioblastoma by gene expression profiling. *Oncogene* 22: 2361-2373, 2003.
- Raza SM, Fuller GN, Rhee CH, Huang S, Hess K, Zhang W and Sawaya R: Identification of necrosis-associated genes in glioblastoma by cdna microarray analysis. *Clin Cancer Res* 10: 212-221, 2004.
- Van den Boom J, Wolter M, Kuick R, *et al.*: Characterization of gene expression profiles associated with glioma progression using oligonucleotide-based microarray analysis and real-time reverse transcription-polymerase chain reaction. *Am J Pathol* 163: 1033-1043, 2003.
- Nutt CL, Mani DR, Betensky RA, *et al.*: Gene expression-based classification of malignant gliomas correlates better with survival than histological classification. *Cancer Res* 63: 1602-1607, 2003.
- Freije WA, Castro-Vargas FE, Fang Z, *et al.*: Gene expression profiling of gliomas strongly predicts survival. *Cancer Res* 64: 6503-6510, 2004.
- Rich JN, Hans C, Jones B, *et al.*: Gene expression profiling and genetic markers in glioblastoma survival. *Cancer Res* 65: 4051-4058, 2005.
- Dong S, Nutt CL, Betensky RA, *et al.*: Histology-based expression profiling yields novel prognostic markers in human glioblastoma. *J Neuropathol Exp Neurol* 64: 948-955, 2005.
- Liang Y, Diehn M, Watson N, *et al.*: Gene expression profiling reveals molecularly and clinically distinct subtypes of glioblastoma multiforme. *Proc Natl Acad Sci USA* 102: 5814-5819, 2005.
- Simon R: Using DNA microarrays for diagnostic and prognostic prediction. *Expert Rev Mol Diagn* 3: 587-595, 2003.
- Eszlinger M, Krohn K, Frenzel R, Kropf S, Tonjes A and Paschke R: Gene expression analysis reveals evidence for inactivation of the TGF-beta signaling cascade in autonomously functioning thyroid nodules. *Oncogene* 23: 795-804, 2004.
- Eisen MB, Spellman PT, Brown PO and Botstein D: Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* 95: 14863-14868, 1998.
- Kaplan E and Meier P: Non-parametric estimation from incomplete observations. *J Am Statist Assoc* 53: 457-481, 1958.
- Frahm C, Haupt C, Weinandy F, Siegel G, Bruehl C and Witte OW: Regulation of gaba transporter mRNA and protein after photothrombotic infarct in rat brain. *J Comp Neurol* 478: 176-188, 2004.
- Iwadate Y, Sakaida T, Hiwasa T, Nagai Y, Ishikura H, Takiguchi M and Yamaura A: Molecular classification and survival prediction in human gliomas based on proteome analysis. *Cancer Res* 64: 2496-2501, 2004.
- Zukiel R, Nowak S, Barciszewska AM, Gawronska I, Keith G and Barciszewska MZ: A simple epigenetic method for the diagnosis and classification of brain tumors. *Mol Cancer Res* 2: 196-202, 2004.

41. Kim AS: Clinical impact of gene expression profiling on oncology diagnosis, prognosis, and treatment. *Comb Chem High Throughput Screen* 7: 183-206, 2004.
42. Oehring RD, Miletic M, Valter MM, Pietsch T, Neumann J, Fimmers R and Schlegel U: Vascular endothelial growth factor (vegf) in astrocytic gliomas - a prognostic factor? *J Neurooncol* 45: 117-125, 1999.
43. Leins A, Riva P, Lindstedt R, Davidoff MS, Mehraein P and Weis S: Expression of tenascin-c in various human brain tumors and its relevance for survival in patients with astrocytoma. *Cancer* 98: 2430-2439, 2003.
44. Rempel SA, Golembieski WA, Ge S, Lemke N, Elisevich K, Mikkelsen T and Gutierrez JA: Sparc: a signal of astrocytic neoplastic transformation and reactive response in human primary and xenograft gliomas. *J Neuropathol Exp Neurol* 57: 1112-1121, 1998.