

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^{1,2}

Sophie Godard, Gad Getz, Mauro Delorenzi, Pierre Farmer, Hiroyuki Kobayashi, Michimasa Nozaki, Annie-Claire Diserens, Marie-France Hamou, Pierre-Yves Dietrich, Luca Regli, Robert C. Janzer, Philipp Bucher, Roger Stupp, Nicolas de Tribolet, Eytan Domany, and Monika E. Hegi³

Lab of Tumor Biology and Genetics [S. G., H. K., M. N., A.-C. D., M.-F. H., N. d. T., M. E. H.] of the Dept of Neurosurgery [L. R., N. d. T.], Multidisciplinary Oncology Center [R. S.], and Div of Neuropathology [R. C. J.], University Hospital (CHUV), 1011 Lausanne, Switzerland; Dept of Physics of Complex Systems, Weizmann Institute of Science, Rehovot 76100, Israel [G. G., E. D.]; Swiss Institute of Bioinformatics (SIB) and Swiss Institute for Experimental Cancer Research (ISREC), 1066 Epalinges, Switzerland [M. D., P. F., P. B.], Department of Oncology, Hôpital Universitaire Genève (HUG), Switzerland [P.-Y. D.]

Running title: Gene expression profiles classify astrocytic gliomas

Key words: glioma, cDNA-array, gene expression profiling, tumor classification, cluster analysis.

¹This project was supported by grants of the Swiss National Science Foundation and ONCOSUISSE (to M. E. H. and N. d. T.), by the National Center of Competence in Research Molecular Oncology (NCCR) (M. E. H., P. F. and M. D.), by the Association des neuro-oncologues d'expression française (ANOCEF) (to S. G.), the Germany-Israel Science Foundation (GIF), the Israel Science Foundation (ISF), Minerva and the Ridgefield Foundation (to E. D. and G. G.).

²Supplementary data for this article to be placed at Cancer Research Online (<http://cancerres.aacrjournals.org>)

³Correspondence : M. E. Hegi, Laboratory of Tumor Biology and Genetics, Department of Neurosurgery, Centre Hospitalier Universitaire Vaudois (CHUV), BH19-110, 1011 Lausanne, Switzerland

email : Monika.Hegi@chuv.hospvd.ch

Phone: +41-(21) 314 2582

Fax: +41-(21) 314 2587

⁴The abbreviations are: LGA, low grade astrocytoma; PrGBM, primary glioblastoma; ScGBM, secondary glioblastoma; EGF, epidermal growth factor; PDGF, platelet derived growth factor; CTWC, Coupled Two-Way Clustering; OAIII, anaplastic oligoastrocytoma; RecGBM, recurrent glioblastoma; CL, cell line; SPC, Super Paramagnetic Clustering; MDS, multidimensional scaling; FDR, false discovery rate; k -NN, k nearest neighbor; VEGF, vascular endothelial growth factor; VEGFR1, vascular endothelial growth factor receptor type 1; PTN,

pleiotrophin; IGFBP2, insulin like growth factor binding protein 2; HIF-1, hypoxia-inducible factor-1; IGF-1, insulin-like growth factor-1; TGF β -2, tumor growth factor β -2; bFGF, basic fibroblast growth factor

⁵The URLs referred to are: CTWC-Server at the Weizmann Institute (<http://ctwc.weizmann.ac.il>); homepage of complete data analysis (<http://www.weizmann.ac.il/physics/complex/compphys/gbm>); The Comprehensive R Archive Network, <http://www.R-project.org>.

ABSTRACT

The development of targeted treatment strategies adapted to individual patients requires identification of the different tumor classes according to their biology and prognosis. We focus here on the molecular aspects underlying these differences, in terms of sets of genes that control pathogenesis of the different subtypes of astrocytic glioma. By performing cDNA-array analysis of 53 patient biopsies, comprising low-grade astrocytoma, secondary glioblastoma (respective recurrent high-grade tumors), and newly diagnosed primary glioblastoma, we demonstrate that human gliomas can be differentiated according to their gene expression. We found that low-grade astrocytoma have the most specific and similar expression profiles, whereas primary glioblastoma exhibit much larger variation between tumors. Secondary glioblastoma are in between and display features of both other groups. We identified several sets of genes with relatively highly correlated expression within groups that a) can be associated with specific biological functions and b) effectively differentiate tumor class. One such gene cluster discriminating primary versus non-primary glioblastoma, comprises mostly genes involved in angiogenesis, including VEGF, VEGFR1, but also IGFBP2, that has not yet been directly linked to angiogenesis. The separating groups of genes were found by the unsupervised Coupled Two-Way Clustering method, and their classification power was validated by a supervised construction of a nearly perfect glioma classifier.

INTRODUCTION

Due to their diffusely infiltrating behavior LGA⁴ (low grade astrocytoma, WHO grade II) cannot be resected completely, and will usually recur. At relapse, progression to anaplastic astrocytoma (WHO grade III) or glioblastoma *multiforme*, the most malignant form of gliomas (WHO grade IV), is common. Most glioblastoma arise *de novo* without evidence of a less malignant precursor lesion, and are termed primary glioblastoma (PrGBM). Glioblastoma evolving from a previous lower grade glioma are defined as secondary glioblastoma (ScGBM). Even though PrGBM are indistinguishable from ScGBM by histology, the two kinds of tumors exhibit distinct genetic alterations and occur in different age groups. The peak incidence for PrGBM is in older people with a mean age of 55 years, while ScGBM are typically diagnosed in younger patients (<45 years). Thus, PrGBM and ScGBM can be considered as two different diseases (1), despite a similarly grim outcome with a median survival of less than one year after diagnosis and no effective therapy. PrGBM are characterized by amplification/rearrangement and overexpression of the EGF (epidermal growth factor) receptor gene (in 40% and 60% of the patients, respectively) often in association with deletion of the INK4a/p14ARF gene locus (2). The hallmarks for ScGBM are p53 mutations (60%) and overexpression of PDGF (platelet derived growth factor) and PDGF receptor (3, 4). Recent advances in the understanding of the molecular bases of these two distinct genetic pathways and their implication for tumor initiation and progression, and regarding their cell of origin, have come from the development of respective mouse glioma models and developmental neurobiology (5). However, a significant fraction of tumors cannot be characterized by either of the two pathways depicted above, suggesting additional pathogenetic pathways that need to be uncovered. Further, current knowledge of tumor genetics has not yet yielded clinically relevant factors predictive for outcome or response to therapy. In view of recent advances in therapeutic approaches to treat cancer by specifically targeting deregulated pathways (6, 7) more detailed knowledge of underlying mechanisms and their relevance for the cancer process is necessary. This will allow rational design of future treatment modalities tailored according to the biology of the individual tumors.

An essential preliminary step towards this goal is the establishment of a taxonomy of tumors on the basis of their gene expression profiles. A search for alternative pathways must be based on identification of genes whose expression differs significantly between the various tumor classes. In particular, it is important to look for a group of (possibly) co-regulated genes, some of which share some known biological function, and whose expression differentiates tumor classes. Identification of such groups leads to better understanding of the biological processes that underlie the distinction between the tumors and also may provide clues for the roles of genes whose role in the initiation and progression of cancer is not yet known.

Here we use gene expression profiling to identify genes whose expression levels reflect biological processes characteristic of distinct subtypes of tumors. In particular, we looked for genes whose expression profiles differentiate three groups of astrocytic glioma: LGA, ScGBM and PrGBM. Standard supervised statistical analysis yields a large number of such genes, and we looked for general features of the expression profiles characteristic of the tumor classes. Then we turned to search for groups of genes whose expression is relatively highly correlated over a subset of the tumors, indicating possible co-regulation. Within each such correlated group we identified genes whose biological function is known. This information is then used to enhance our understanding of malignant progression and differences between tumor classes on the one hand, and for discovery of gene function, by providing

clues for the yet unknown biological function of other members of a correlated gene cluster, providing potential molecular targets for therapy.

The search for such correlated groups of genes, that distinguish between various classes of tumors, has been accomplished by Coupled Two-Way Clustering (CTWC) (8). This unsupervised method has been applied with success to analyze expression data from colon cancer, leukemia, breast and skin cancer (8-10). Finally, we combine unsupervised with supervised analysis in order to validate the separating power of the gene clusters that were identified, by constructing a tumor classifier that discriminates between LGA, ScGBM and PrGBM.

MATERIALS AND METHODS

Tumor biopsies and cell lines. Tumor biopsies, obtained from patients who underwent surgery at University Hospitals in Lausanne or Geneva, or the Cantonal Hospital in Fribourg (Switzerland), were shock frozen and stored at -70°C . The use of biopsies and respective clinical data has been approved by the local ethics committee and the respective federal agency. The tumors were diagnosed according to the WHO classification 2000 (1). Twenty-one biopsies originated from twenty patients enrolled in a prospective pilot trial for newly diagnosed glioblastoma (11). These biopsies are called hereafter primary glioblastoma (PrGBM), with the exception of three biopsies: one recurrent tumor 1497 (of PrGBM 1430, RecGBM), biopsy 1342 that was a ScGBM, it had progressed from a low grade astrocytoma (WHO grade II; LGA) that was surgically removed 13 years before, but not treated otherwise, and biopsy 1357 that was revised as a mixed anaplastic oligoastrocytoma (WHO grade III, OAI) upon central review. From the tumor bank additional 32 gliomas from twenty-two patients were included, comprising 24 LGA, two of which were recurrent LGA, and 8 respective high grade recurrent tumors called ScGBM (comprising two astrocytoma WHO grade III and 7 grade IV). These tumors have been previously analyzed for p53 mutations (12). Additional information on the individual tumors, including p53 mutation, age and gender of the patients is available as supplementary information² (Table S1). The tumor samples were organized into two data sets according to their date of analysis: the first, subsequently used as a training set, comprising 14 PrGBM, 5 ScGBM, and 12 LGA; and the second, used as validation set, 4 PrGBM, 4 ScGBM and 12 LGA. Normal brain tissue for RNA isolation was obtained from a lobectomy after brain edema, and additional samples of human normal brain total RNA was obtained from Clontech (human total RNA panel IV). Culture conditions for the glioblastoma cell line U87, LN-229, LN-Z308, and the p53-inducible glioblastoma cell line 2024, which is derived from LN-Z308 after introducing the Tet-On System, has been described before (13). Expression of wild-type p53 was induced with $2\mu\text{g/ml}$ doxycycline in cell line 2024 during 24 hours before RNA-isolation (2024+). For anoxia treatment, cells were cultured for 20 h in an anaerobic culture incubator ($\text{O}_2 < 1\%$) (Scholzen, Microbiology Syst AG) filled with a mixture of N_2 , H_2 and CO_2 .

Isolation of total RNA. Prior to RNA isolation a section of the frozen tumor biopsy was reevaluated after hematoxyline / eosine staining by the neuropathologist (R. C. J.) to estimate the proportion of solid tumor, contaminating normal tissue, or infiltration zone. Pieces comprising more than 30% normal tissue were excluded. Fifty to 100 mg of frozen tumor tissue were homogenized in Trizol solution (Life Technology). RNA phase was

purified in saturated phenol solution (60% phenol, 15% glycerol, 0.1M sodium acetate, pH 4). The quality of the RNA was evaluated on agarose gel. RNA from cell lines was isolated similarly.

p53-mutation analysis. The tumors were screened for p53 mutations using the yeast functional assay as described (14-16), followed by direct sequencing, if the test was positive (Microsynth, Balgach, Switzerland).

Northern blot analysis. Northern blot analysis was performed as described before using 10 µg of total RNA (17). The membrane was sequentially hybridized to the plasmid derived probe VEGF165 (EcoRI/BamHI-fragment of pBspI-KS-VEGF165 (18), and the PCR derived probes for IGFBP2, 3 & 5 (respective primer sequences provided by Clontech). Probes were radioactively labeled employing the random primed DNA labeling kit (Boehringer) using [α - 32 P]dCTP (3000 Ci/mmol, Amersham). Expression was quantified by phosphorimager (Fuji, BAS 1000).

cDNA synthesis and hybridization. Atlas Human Cancer 1.2 Array membranes (Clontech) were used for all experiments described. These nylon filters are spotted with 1185 genes, including reference genes and 1176 genes related to cancer. Three to 5 µg of DNase treated total RNA was used to prepare a labeled 1st strand cDNA using the Clontech kit, basically as recommended. Briefly, RNA in a volume of 2 µl was mixed with 1 µl of specific CDS primers, 1 µl of RNasin (Promega, 40u/ml), and denatured at 70°C. Subsequently, 1 µl of Superscript II reverse transcriptase (Gibco BRL, 200u/ml) and 3.5 to 5 µl of α - 32 P-dATP (3000 Ci/mmol, Amersham) were added, and the reaction was performed at 48°C for 30 min. The probe was purified with Clontech Atlas Nucleospin extraction kit following the manufacturer's recommendations. Prior to use the membranes were boiled in 0.5% SDS solution for stripping and also at first use. Membranes were exposed to an imaging plate (BAS MS 2040, Fuji) for 1 to 8 days. The Atlas Cancer Arrays were used three times.

Preprocessing and analysis of expression data. Expression was quantified by phosphorimager (Fuji BAS1000) and analyzed with Atlasimage 1.5 software (Clontech). After background subtraction signals were normalized using the "sum method" comparing the sum of the intensity of all genes in the experiment to the sum calculated from the reference sample, yielding the coefficient of normalization, "c". The reference utilized in all experiments represents the calculated average of five independent expression profiles derived from "normal" brain. The ratio (R) was calculated for every gene as follows:

$R = [(s - s_0) * c + K] / [(r - r_0) + K]$, where $K = b * [s_0 + (c * r_0)]$; s, gene intensity in sample; s₀, background in sample; c, coefficient of normalization obtained using the sum method; r, gene intensity in reference experiment; r₀, background in reference experiment; and b=2. For $s \gg s_0$ and $r \gg r_0$ this formula reduces to the normalized ratio of sample and reference. In other cases it circumvents the problem of losing valuable information if the experiment or the reference display no expression for a given gene (division by 0), and generates ratios R shrunk towards 1 for low expressed genes in both experiments. Furthermore, introduction of "K" filters out data derived from genes marginally expressed over background by decreasing their ratios R.

Distance and Multidimensional Scaling (MDS). Distances between samples were given by the Euclidian distances in the space of logarithms of the ratios R. Multidimensional Scaling was performed with the implementation for the classical metric scaling (also known as principal coordinate analysis) (19) available in the mva package for R available online at the Comprehensive R Archive Network.⁵

Coupled Two-Way Clustering (CTWC). A variation filter was applied; only those genes were kept, for which the ratio of the maximal and minimal R values (obtained for the 36 experiments) exceeded 2. For each gene (row), the log of the ratio R was mean-centered (subtracting the average) and normalized, and the results were stored in the expression matrix. Euclidean distances were measured between all pairs of genes and between all pairs of tumors, which serve as the input to our clustering procedure. CTWC has been described elsewhere, see Getz *et al.* (8) for full details and comparisons with other methods; (9, 10) for its applications to colon, breast, and skin cancer data analysis; and (20) describing the use of the publicly available CTWC – Server⁵. CTWC is an iterative process; the first steps are standard, clustering all genes on the basis of the data from all tumors, and clustering all tumors, using data from all genes. In both resulting dendrograms we identify stable (i.e. statistically significant, see below) gene and sample clusters; these are denoted, respectively, as GX or SY (where X,Y are running indices). Note that G1 represents the set of all genes, and S1 that of all samples. In the second iterative step, each *stable* gene cluster GX is used to characterize and cluster the members of every sample cluster SY, and vice versa.

In order to use CTWC we must be able to identify stable clusters. One of the few algorithms that provide a stability index to each cluster is *Super Paramagnetic Clustering (SPC)*, a physics-based method, that has been described in full detail in (21), tested on a large number of problem areas, ranging from image analysis and speech recognition to gene expression (21-23).

The parameter T controls the resolution at which the data are viewed; as T increases, clusters break up. A cluster C is “born” at $T=T_1(C)$, the value of T at which its “parent” cluster breaks up into two or more subclusters, one of which is C . As T increases further, to $T_2(C) > T_1(C)$, C itself breaks up and “dies”. We consider a cluster “broken up” once it loses more than a certain number (for genes: 3) of its members when T increases by a single step. *SPC* provides a quantitative stability index, $Stab(C) = T_2(C) - T_1(C)$ for any cluster C . The larger $Stab(C)$, the more statistically significant and stable (against noise in the data and fluctuations) is the cluster C (8, 24). We register C as stable only if it exceeds a certain size and when $Stab(C)$ exceeds a certain threshold.

It should be noted that standard hierarchical clustering techniques such as average linkage did not succeed to partition the tumors according to their clinical labels.

Classification. For binary class comparisons two standard statistical tests were used: (i) a parametric test, the two-sample t-test (with unknown but equal variances) testing the null hypothesis that the mean expression levels of the two classes are equal and (2) a non-parametric test, the Wilcoxon ranksum test that tests whether the two populations are identical. In order to address contamination with false positive genes associated with *multiple comparisons* we use the method of Benjamini and Hochberg (25) that bounds the average false discovery rate (FDR): namely, the fraction of false positives among the list of differentiating genes. To bound the FDR by q , one orders the N genes according to their P-values, $P_{(1)} \leq P_{(2)} \leq \dots \leq P_{(N)}$ and rejects the null hypothesis for those genes whose index, i , is less or equal to $i^* = \max_j \dot{P}_{(j)} \leq j \times q/N$. The outcome of this method is a list of genes, of which at most a fraction q are false positives.

The class discrimination power of the selected sets of genes was further validated by applying the respective k-nearest neighbor (k -NN) classifier constructed on the training set of the tumor tissue samples to predict the class of an independent validation set of 20 additional samples. Computations were performed with the class package for R

and with $k=3$. The metric used was Euclidian distance in the space of the ratios R, classification was decided by majority vote, with ties broken at random. When ties occurred, classification was repeated 100 times and p-values were averaged. P-values for the significance of the deviation from independence between true and predicted labels were obtained with the Fisher exact test (ctest package for R) and with the alternative hypothesis set to "greater" in the 2 by 2 case.

RESULTS

Gene expression profiles separate tumor classes. In order to characterize and classify gliomas by their gene expression profiles, RNA isolated from frozen gliomas and 3 experiments with glioblastoma cell lines was analyzed using cDNA-arrays comprising 1185 genes. The configuration of the 51 astrocytic gliomas in Euclidian space of overall gene expression is visualized in Fig. 1 using multidimensional scaling (MDS) that attempts best preservation of distances between the points in the original multidimensional space. MDS analysis clearly suggests that the gene expression profiles contain information discriminating the three classes of astrocytic gliomas. The two classes whose points show a higher degree of spatial intermingling are the LGA and the ScGBM, while the best separation is between LGA and PrGBM. This correctly reflects other biological characteristics of the ScGBM, in that they progress from LGA, but share the malignancy grade (WHO IV) with PrGBM, and are indistinguishable from them histopathologically. Interestingly, the pairwise distances in space of all genes (1185) are highest among PrGBM (mean, 10.06; stdev, 2.10) and shortest between LGA (mean, 6.09; stdev, 1.18), while ScGBM are intermediate (mean, 7.27; stdev, 1.18). This suggests more biological heterogeneity of PrGBM, without evidence of obvious subclasses, reflecting well the heterogeneous morphology characteristic for glioblastoma *multiforme*. This may indicate a challenge to find outcome predictors and suitable targets for future therapies.

Supervised Analysis. We applied t-test and the Wilcoxon ranksum test to look for genes that differentiate two known classes, using all 1185 genes and the training-set of 31 astrocytic gliomas. The inherent problem with this approach, that of multiple comparisons, is overcome by the False Discovery Rate (FDR) method (25). The threshold for FDR was set at $q=0.05$, which supplies a list of separating genes with the expected number of false positives kept at 5%. First we looked for genes whose expression levels separate the 14 PrGBM samples from the combined ScGBM & LGA (5 + 12) and found 191 discriminating genes using t-test and 174 with ranksum; the union of the two lists contained 205 genes (shared genes 160), of which only 10 are expected to be falsely included (the highest p-value within the 205 genes is 0.0078). The same tests yielded 126 and 98 genes, respectively, separating PrGBM & ScGBM from LGA; the union of these two lists is 132 genes (92 shared genes; highest p-value of 0.0051). Similarly, 187 genes separated PrGBM from LGA (union of 167 and 163 separating genes (143 shared genes, highest p-value – 0.0067). Full results of these supervised analyses, including lists of selected genes of all binary comparisons, can be found in Table S2 of the supplementary information. Note that the lists of genes determined by using the t-test and Wilcoxon ranksum test are very similar. Due to the low number of the ScGBM in the training set, comparisons involving ScGBM, ScGBM vs PrGBM and ScGBM vs LGA, yielded only very few separating genes (1 and 3, respectively).

Our efforts to find genes correlated with p53 status failed, whether we used all tumors, or only subsets thereof. Neither did we find any gene cluster separating groups of samples on the basis of their p53 status (see p53

status of biopsies in Table S1 of supplementary information). Our results are in line with a previous report on a series of LGA, the authors neither found evidence for genes correlated with p53 status (26). This result might not be surprising in consideration of the fact that the p53 pathway has been found to be inactivated, one way or another, in over 70% of all astrocytic gliomas, regardless of tumor grade, by mutational alteration of p53 or alternative mechanisms (27). Two other mechanisms are overexpression of mdm2, a negative regulator of p53, or inactivation of p14ARF, a negative regulator of mdm2, by deletion or promoter methylation.

Unsupervised analysis: Coupled Two Way Clustering (CTWC). Unsupervised methods have to be used for search and discovery of previously unpredicted correlations, structure and partitions in the data. CTWC was used in order to search for genes that form a group with correlated patterns of expression and induce clear separation of the samples into classes. **First level CTWC:** The data derived from 36 experiments, comprising the 31 astrocytic gliomas of the training set, two additional gliomas (a RecGBM and a OAI), and the 3 experiments with glioblastoma cell lines, were subjected to a variation filter reducing the number of genes (rows) to 358. The first step of the analysis is to cluster the set of all 358 genes (G1), using the data from the set of all 36 experiments (S1), and to cluster the tumors S1 using their expression profiles measured for all the genes G1. The first clustering operation is denoted as G1(S1) and the second as S1(G1). The first produces a dendrogram of gene clusters, and the second – one of tumors. These two dendrograms are shown in Fig. 2A together with the correspondingly reordered expression matrix. The results of these two initial steps are summarized as follows: Clustering the genes using all experiments G1(S1) identified 15 stable gene clusters (*Stability* > 8, corresponding to $P < 0.01$) that are indicated in the gene dendrogram of Fig. 2A, and denoted G2 to G16. The operation S1(G1), of clustering the tumors using all genes, yields a dendrogram with two stable clusters, S2 and S3. Details of the clusters obtained in the CTWC can be viewed and searched online at the respective webpage at the Weizmann Institute⁵.

Second level CTWC: clustering tumors using selected gene clusters. The first goal of our analysis was to look for clusters of genes that partition the tumors into groups that reflect their classification as glioblastoma versus LGA. Since “misclassification” is possible (see below), an unsupervised search to discover new patterns is suitable. We use all the stable gene clusters found above, one by one, to recluster S1, look for stable partitions and when one is found, we check whether it divides the tumors, to a good approximation, into glioblastoma or LGA. In general, every stable partition of the tumors is searched for some biologically relevant attribute.

Separation into PrGBM versus LGA and ScGBM based on angiogenic activity: S1(G5). In the first level of CTWC we identified G5 as a stable cluster of 9 genes (Figs. 2A & B). This cluster comprises hallmarks of angiogenesis such as VEGF (vascular endothelial growth factor), VEGFR1 (vascular endothelial growth factor receptor type 1), and PTN (pleiotrophin) (28-30), and some of the other genes have been related to angiogenesis before. When the expression levels of only these genes are used to characterize the tumors, a large and stable cluster, S11, of 21 tumors emerges (Fig. 2B). This cluster contains all the 12 LGA and 4 of 5 ScGBM. Of the remaining 5 samples of S11, three are cell lines, whose expression profiles are consistently different. Hence the expression levels of G5 give rise to a nearly perfect separation of clinically defined tumor entities, namely PrGBM from LGA and ScGBM. These genes are significantly upregulated in PrGBM (see Fig. 2B) as compared to LGA and ScGBM.

Interestingly, a compilation of the literature provides evidence that apart of the angiogenic hallmarks such as VEGF, some additional genes of this cluster are associated with aspects of angiogenesis: IGF2BP2 (insulin like

growth factor binding protein 2) has been suggested to be activated by HIF-1 (hypoxia-inducible factor-1), although, in absence of a defined HRE (hypoxia-response element) containing a putative HIF-1 binding site it may also be an indirect effect (31). c-Jun and TIS11B are early transcription factors to mitotic signaling, and take part in the EGF/MAP-kinase pathway. Recently, c-Jun has been shown to functionally cooperate with HIF-1 in hypoxia-induced gene transcription (32). TIS11B can be induced in response to EGF (epidermal growth factor) but also IGF-1 (insulin-like growth factor-I) stimulation (33). A role for IGF-I in angiogenesis has been established (31, 34, 35). Gravin is a protein of the AKAP (A-Kinase Anchoring Protein) family and may play a role in the modulation of cell motility and adhesion (36), and has been reported to be upregulated in cultured endothelial cells under conditions suggestive of a role in wound repair and vascular development (37). Other genes in this cluster are SOCS-2 and -3 that in general were found to be downregulated in most gliomas as compared to normal brain, but more so in non-PrGBM, as visualized in Fig. 2B. SOCS are a class of proteins that are negative regulators of cytokine receptor signaling via the JAK/ STAT pathway, reviewed in (38). SOCS-2 has tumor suppressing function, and respective knock-out mice exhibit a giant phenotype due to unopposed growth, similar to IGF-I or GH (growth hormone) transgenic mice. SOCS-3 may not inhibit all signaling, it can also bind RasGAP and consequently enhance Ras signaling. SOCS-2 and SOCS-3, were demonstrated to be direct binding partners of the IGF1 receptor (39).

Separation by immune response related genes; S1(G12). The gene cluster G12 contains 13 genes that are mainly related to the immune system and to inflammation. The dendrogram obtained by clustering the tumors using these genes is presented in Fig. 2C. Of its two large stable clusters, S39 contains 9 tumors, 8 of which are PrGBM; the expression level of the G12 genes in these tumors is high. On the other hand, S38 has all but one of the LGA and ScGBM samples, together with 4 PrGBM; 1316, 1284, 1399 and 1297. Three of these four tumors are from patients whose survival was relatively long (> 18 months) (11). Hence the immune system related genes of G12 appear to have lower expression levels in LGA and ScGBM, and in those PrGBM that have longer survival.

Three genes comprised in cluster G12, CD16, CD64, and FcRn are from the same family of Fc gamma receptors. DAP12 and some of the Fc gamma receptors are markers of NK (natural killer) cells, although their expression is not exclusive. Four of the genes, CD64, CYBB, Leu13 (CD225), and interferon-regulated resistance GTP-binding gene (MXA), are regulated by interferon. And expression of allograft inflammatory factor-1 (AIF-1) has previously been shown to define a subset of infiltrating macrophages and microglial cells in gliomas (40). Glutathione-S-transferase omega (GST-omega) expression is ubiquitous, and was also found in glial cells and some types of macrophages (41). This gene family plays an important role in detoxification and response to oxidative injury. Increased expression of the genes in G12 in samples separated in S39 in general was found in accordance with a more prominent presence of CD68 (macrophages, activated microglia) and CD3 (T-cells) positive cells as determined on respective paraffin sections of the samples (not shown), in line with the role of these genes in inflammation. Studies trying to relate immune response in gliomas with survival have not yielded conclusive results. It is known that gliomas produce an immuno-compromised environment by secreting cytokines such as TGF β -2 (tumor growth factor β -2) (42, 43), which in this study emerged in node N566 that separates glioblastoma from LGA.

Gene selection by combining supervised analysis with clustering. In a second step we used the results of

the supervised analysis described above, to identify in the dendrogram yielded by G1(S1) those clusters that are rich in discriminating genes. Of the 205 genes that were found to differentiate PrGBM from ScGBM & LGA, 91 passed the variation filter and were included in G1. As depicted in Fig. 3A most separating genes belong to one of four marked clusters. The same clusters are identified also as rich in genes that partition samples according to the other supervised binary comparisons mentioned above. Two of the identified clusters are indeed G5 (6 out of 9 genes are among the 91 separating ones) and G12 (5 out of 13). However, two additional clusters are identified as rich in separating genes – these are marked in Fig. 3A as nodes N505 (27 out of 29) and N566 (6 out of 7). These sets of genes were not selected as stable clusters by CTWC since they are either too small or they “leak” too many genes as the control parameter T is increased. However, as clearly seen in the distance matrix (Fig. 3B), the genes of both of these nodes have relatively highly correlated expression profiles. These additional clusters contain some interesting genes: node N505 – members of the Wnt- and Notch-pathway, β -catenin, protein phosphatase 2A B56 α , and notch 2, and apoptosis related genes CAS (cellular apoptosis susceptibility protein) and DAPK1 (death-associated protein kinase) (Fig. 3C); node N566 - TGF β 2, IGFBP3, and chitinase (Fig. 3D). Consistently we found differential expression of different members of the Notch- and Wnt-signaling pathways in LGA as displayed in node N505 (Fig. 3C) and in a number of stable gene clusters partitioning LGA (data not shown, see webpage⁵ for complete unsupervised analysis) suggesting that deregulation of the Notch- and/or Wnt-signaling pathways may play a role in the development of LGA and their malignant progression.

Next, we used these two additional clusters to partition all the samples S1. As shown in Fig. 3C, node N505 gives rise to a nearly perfect separation of the samples into 3 clusters: (i) “Non-PrGBM”; of 17 samples, comprising 12 of 12 LGA, 3 of 5 ScGBM, and 1 PrGBM and the OAI, both of which (1308 and 1357) joined the non-PrGBM cluster in S1(G5). (ii) mainly PrGBM - 13 out of 14 PrGBM, the RecGBM (1497), 2 of 5 ScGBM, one of which (1342) had joined the PrGBM in S1(G5), 1 cell line (U87), and (iii) a cluster with the remaining cell line under two conditions (wild-type p53 on and off; 2024+, 2024-). Hence node N505 gives a clear separation into LGA vs. PrGBM, with the ScGBM split evenly.

On the other hand, as seen in Fig. 3D, node N566 identifies a cluster of 20 samples comprising 12 of 12 LGA, thus, this node separates the samples largely into LGA vs. PrGBM, with the ScGBM distributed evenly.

Validation of the results. A tumor classifier was built in order to validate our findings. A successful test of such a classifier based on the selected genes and applied to an independent set of samples indicates that the separation found previously is robust and was not caused by random fluctuations. The validation set consisted of 20 samples: 4 PrGBM, 4 ScGBM, and 12 LGA. For each separating cluster of genes (G5, G12, node N505 and node N566) and their combination we generated a k -Nearest Neighbor classifier ($k=3$) and tested its ability to perform class partitions. In all cases, the separation power was found to be significant (see Table 1). In addition we tested whether a three-way separation of the tumor types can be performed based on the combination of the 4 genes clusters. This gave rise to only 1 to 2 errors out of the 20 gliomas, thus validating our claim that the tumor types have different gene expression profiles that can be used to distinguish between them ($p=2.49E-06$). When using the gene sets separately, the distinction of PrGBM versus ScGBM is the least successful. Most classification errors disappear when all four selected clusters are used. This suggests that distinct sets discriminate between different aspects of the

classes. In fact, genes of the same cluster are generally highly correlated across the samples of the same class as a result of the selection procedure, while genes of different clusters are not. Although the best performance is obtained by combining all genes, the G5 set alone provides already a good separation power in all cases, including 3-way classification where only 1 to 2 (8%) errors were made (Table 1). Most of the genes in G5 are involved in angiogenesis in a broader sense as detailed above. Differential expression of two prominent members of this cluster, VEGF and IGFBP2, according to the 3 tumor subtypes is visualized in a box-plot in Fig. 4.

The tumor classifier using expression data of all four identified discriminatory clusters yielded correct tumor prediction in 93% of an independent set of astrocytic gliomas. This remarkable success has to be appreciated in light of the inherent inaccuracy of the initial tumor labels for gliomas, due to sampling errors and subjective histological criteria that lead to striking interobserver differences in the range of 70% among any two neuropathologists (44). In this context four tumors originally designated as PrGBM attracted our attention (1342, 1297, 1308, 1357), they appear evenly in clusters characteristic for PrGBM and LGA, respectively, depending on the analysis (Figs. 2 and 3). The first turned out to be a ScGBM (1342), progressed from a LGA surgically removed some 13 years ago (histologic diagnosis unchanged; PrGBM versus ScGBM anamnestic criteria). Central pathology review revised the second (1357) as OAIII, while the other two cases were confirmed. Tumor 1297, clinically defined as a PrGBM, presented with genetic characteristics of a ScGBM, namely p53 mutation and most prominent overexpression of PDGFR. Thus, the gene expression profile of 1297 supported the impression already indicated by molecular genetics that this tumor is not a “classic” PrGBM. The notion that a fraction of PrGBM, as defined by clinical criteria (astrocytoma grade IV at first diagnosis, no clinical evidence for a lower grade precursor lesion), effectively are unrecognized ScGBM, has been brought forward before (45). Finally, glioblastoma 1308 exhibited overexpression of EGFR, characteristic for PrGBM, but the biopsy utilized for gene expression profiling comprised a part of the tumor displaying only features of WHO grade II, despite the fact that the respective histological diagnosis of the tumor was reconfirmed as WHO grade IV on the diagnostic material (the most malignant part of tumor defines diagnosis). This is a very interesting case, it demonstrates that regions of a glioblastoma with the appearance of a lower grade astrocytoma have already features only found in PrGBM, exemplified here as overexpression of the EGFR gene and the genes comprised in G12 and N566, characteristic for PrGBM. In contrast, after reclustering S1 (all samples) according to G5, comprising genes indicating high angiogenic activity, this tumor (PrGBM 1308) partitions with LGA. This is biologically convincing, because lower grade astrocytoma have no necrosis and thus are expected to suffer from less hypoxic stress. This observation might be of crucial clinical relevance. Due to the sampling problem during surgery that is aggravated in stereotactic biopsies, the true tumor grade may be underestimated. The identification of markers specific for glioblastoma, however, that may be also detectable in apparent lower grade areas of such a tumor would allow to improve diagnostic reliability, and ultimately the choice of therapy.

Anoxia induces expression of IGFBPs in glioblastoma cell lines. An important question to ask is whether cluster methods can predict some unexpected new gene function based on the assumption that co-expressed genes might be co-regulated and therefore might be involved in the same biological process. IGFBP2 expression displayed a high correlation ($r=0.85$) to VEGF expression; both are overexpressed in PrGBM, as seen Fig. 2B and 4. This raised the question whether IGFBP2 participates in the same biological process as VEGF in these tumors. Indeed, in our

experiment glioblastoma cell lines LN308 and LN229 displayed IGFBP2 upregulation concomitant with VEGF after 20 h of anoxia (Fig. 5). In line with these observations, immunohistochemical findings had revealed that presence of IGFBP2 protein was highest in the vicinity of necrosis in glioblastoma (46). Interestingly, in PrGBM increased expression of IGFBP2 was often paralleled by overexpression of either IGFBP3 (comprised in N566, Fig. 3D) or IGFBP5, or both. Thus, these two family members might also be induced in the same process in these tumors. This is in line with our experiment in glioblastoma cell lines that demonstrated induction of at least one of these three family members under anoxia (Fig. 5). Induction of different members of the family in these cell lines is in line with a certain redundancy of these genes observed in IGFBP-2 knock-out mice. They show increased levels of other IGFBPs, and display a much less dramatic phenotype than the one initially predicted based on the fetal IGFBP-2 expression pattern (47).

DISCUSSION

Our characterization of astrocytic gliomas by their gene expression profiles revealed consistent inherent differences between LGA, ScGBM, and PrGBM. Thus - the previous histological and clinical recognition of these three glioma entities can be strongly supported by cDNA-array data as reflected by MDS of overall gene expression (Fig. 1). The LGA were found to be the most closely related group, and well separated from PrGBM that displayed the most heterogeneous expression profiles. The ScGBM share some of the features of both subgroups, reflecting the common pathogenetic pathway with LGA and the malignancy with PrGBM.

Analysis of the expression profiles for correlated genes separating tumor subtypes allowed identification of sets of genes implicating particular biological features that are associated with tumor specific subtypes. Most strikingly, a cluster (G5) of correlated genes suggests that inherent angiogenic activity, manifested in overexpression of known angiogenic factors including the most potent, VEGF, and VEGFR1, and PTN, distinguishes PrGBM from the other two groups. In fact, this feature alone was sufficient to classify most (92%) astrocytic gliomas correctly according to their subtype. PrGBM appear to have higher angiogenic activity mediated by these factors than LGA, but also higher activity than ScGBM. This difference between PrGBM and ScGBM is somewhat unexpected, since they have the same malignancy grade (WHO IV) and cannot be discriminated by classical histology. This may reflect the fact that PrGBM are rapidly growing tumors, which cannot keep up with their increasing need for blood supply, and thus may suffer from more severe hypoxic conditions triggering angiogenic activity. In contrast, ScGBM may progress over years from LGA, and thus may employ different pathways for sustained angiogenesis (e.g. LGA display higher expression of the angiogenic factor bFGF (basic fibroblast growth factor) (N505); all LGA vs. all PrGBM, $p=0.0099$). This biologic difference may have important implications for response to anti-angiogenic therapy, which is about to enter the clinics (48).

Angiogenesis, an essential requisite of tumor growth, is a tightly controlled, complex process involving various pro- and anti-angiogenic factors that need to act in a concerted fashion in order to yield functional blood vessels (49). Vascular specific growth factors and their respective receptors are VEGF and VEGFR 1 & 2, the angiopoietins with their Tie receptors, and Ephrins together with their Ephrin receptors. Other angiogenic factors, expressed abundantly in gliomas, also act on other cell types, and include bFGF, PDGF, and TGF- β (50).

Glioblastoma are among the most vascularized tumors with characteristic, structurally and functionally abrogated blood vessels of multilayered glomeruloid pattern. Recently, the relative abundance of two vascular types in glioblastoma, described as “classic” pattern with evenly distributed capillary-like microvessels, as opposed to unevenly distributed glomeruloid bizarre vascular formations, have been associated with clinical outcome (51). It remains to be seen how PrGBM and ScGBM differ in this regard.

An important implication of the CTWC method concerns investigation of the genes that belong to identified clusters, here in particular G5, for their biological function. The fact that IGFBP2 is strongly correlated with the angiogenesis genes came as a surprise that will be worth detailed further studies. If one assumes that co-expressed genes might be co-regulated and therefore might be involved in the same biological process, the presented association provides some evidence that IGFBP2 overexpression in astrocytic gliomas may be implicated in angiogenesis. This correlation adds to the fact the IGFBP2 can be induced under hypoxic conditions in glioblastoma cell lines (Fig.5) and mouse embryonic stem cells (31).

Gene expression profiling has discovered IGFBP2 overexpression in several tumor types, such as breast cancer (52), prostate cancer (53), but also most malignant glial tumors (26, 54-56). Only recently, overexpression of members of the IGFBP family, in particular IGFBP2 (reviewed in (57)), has been associated with tumorigenesis in humans, supported by some mouse models (58). The IGFBPs were first identified and characterized based on their role to bind and modulate the *in vivo* bioactivity of the mitogenic and anabolic peptides IGF-I and IGF-II. More recently it has become clear that IGFBPs are multifunctional proteins with IGF-dependent and independent functions in controlling growth and metabolism. The intrinsic bioactivity of IGFBPs, in a positive or negative fashion, depends not only on the cell type but also on their differentiation state, and physiologic/pathophysiologic condition (57).

Our observations support the idea to develop a therapeutic approach targeting the IGF-I / IGFBP system in cancer, as currently discussed.

Another separating gene cluster emerging from CTWC analysis associated with known biological function is G12. It comprises only genes related to the immune system, half of which are interferon inducible, and likely are expressed by lymphoid cells and /or macrophages, although some of them are potentially expressed by tumor cells. The cluster partitioned a subgroup of PrGBM with increased expression from LGA. Differential expression of genes related to the immune system has been reported before (59).

Using the gene expression data of four highly informative gene clusters emerging from CTWC analysis, a tumor classifier was constructed using *k*-NN that successfully predicted most of 20 new biopsies (93%). This is a remarkable result in view of the apparent difficulties of objective classification according to WHO criteria.

This novel approach of gene selection, combined unsupervised with supervised analysis to identify clusters rich in genes informative (by supervised analysis) for tumor classification. CTWC was particularly suitable for this task, since it is designed to go beyond clustering of all genes on the basis of the data from all tumors, and clustering of all tumors, using data from all genes. This is of importance because most of the genes for which expression levels

have been measured are *irrelevant* for the partition sought. CTWC proposes identification of correlated groups (clusters) of genes, and using only data from one such group at a time to re-cluster the tumors, or vice versa.

The procedure followed in this paper for selection of discriminatory genes has some interesting features, and allowed us (i) to reduce the large number of discriminatory genes obtained by binary class comparisons, (ii) enabled an almost correct classification of tumor entities, namely discrimination of LGA from ScGBM, and ScGBM from PrGBM, although pairwise comparisons yielded only few separating genes (partially due to small numbers of ScGBM in the training set, $n=5$; threshold of FDR of $q < 0.05$); (iii) by using the signal of a group of correlated genes, the noise of the individual measurements averages out and is reduced; and most important (iv) the identified gene clusters, yielded information on the biological context of co-expressed and possibly co-regulated genes. Such insight may give some indication on biological processes determining tumor entities. This fact is highlighted in this study, in particular by identification of G5, with high discriminatory power on its own, featuring angiogenesis related genes.

Taken together, gene expression profiling of astrocytic gliomas has yielded important information for tumor classification and its underlying molecular bases. We were able to discriminate PrGBM from ScGBM that are indistinguishable by means of histology, by their difference in expression of angiogenic factors. This angiogenic feature may be of importance for the choice of therapy, anti-angiogenic or not, as they may respond differently. This approach is a first step towards molecular diagnostics for astrocytic gliomas that may improve tumor diagnoses in the future, by adding objective criteria. Further, gene expression profiles may improve diagnosis of small biopsies, which have an inherent risk not to be representative of the whole tumor, by identifying characteristic features of the respective subtypes, in particular markers of high malignancy in apparent low grade tumors. Last, but not least, gene expression profiles may ultimately provide a tool for the identification of patients who are most likely to benefit from targeted therapy. It follows that such approaches for outcome prediction need to be established in clinical trials which in turn will allow rational design of future therapies.

ACKNOWLEDGEMENTS

We are indebted to our colleagues who made this study possible by participating actively in the clinical trial and providing primary tumor samples, namely Drs. J.-G. Villemure, F. Porchet, O. Vernet, P. Otten, A. Reverdin, and B. Rilliet. We thank S. Ostermann Kraljevic, and Drs. M. Albertoni, and G. Pizzolato for their collaboration, and acknowledge Drs. P. Walker and I. Desbaillets for critical reading of the manuscript.

REFERENCES

1. Kleihues, P. and Cavenee, W. K. Pathology & Genetics. Tumours of the nervous system. Lyon: IARC Press, 2000.
2. Hegi, M. E., zur Hausen, A., Ruedi, D., Malin, G., and Kleihues, P. Hemizygous or homozygous deletion of the chromosomal region containing the p16INK4a gene is associated with amplification of the EGF receptor gene in glioblastomas, *Int. J. Cancer*. 73: 57-63, 1997.
3. Watanabe, K., Tachibana, O., Sato, K., Yonekawa, Y., Kleihues, P., and Ohgaki, H. Overexpression of the EGF receptor and p53 mutations are mutually exclusive in the evolution of primary and secondary glioblastomas, *Brain Pathol*. 6: 217-224, 1996.
4. Hermanson, M., Funa, K., Koopmann, J., Maintz, D., Waha, A., Westermarck, B., Heldin, C.-H., Wiestler, O. D., Louis, D. N., von Deimling, A., and Nistér, M. Association of loss of heterozygosity on chromosome 17p with high platelet-derived growth factor α receptor expression in human malignant gliomas, *Cancer Res*. 56: 164-171, 1996.
5. Zhu, Y. and Parada, L. F. The molecular and genetic basis of neurological tumours, *Nat Rev Cancer*. 2: 616-626., 2002.
6. Mendelsohn, J. and Baselga, J. The EGF receptor family as targets for cancer therapy, *Oncogene*. 19: 6550-6565., 2000.
7. Shawver, L. K., Slamon, D., and Ullrich, A. Smart drugs: tyrosine kinase inhibitors in cancer therapy, *Cancer Cell*. 1: 117-23., 2002.
8. Getz, G., Levine, E., and Domany, E. Coupled two-way clustering analysis of gene microarray data, *Proc Natl Acad Sci U S A*. 97: 12079-12084, 2000.
9. Getz, G., Gal, H., Kela, I., Notterman, D. A., and Domany, E. Coupled two-way clustering analysis of breast cancer and colon cancer gene expression data, *Bioinformatics* in press, 2003.
10. Dazard, J.-E., Gal, H., Amariglio, N., Rechavi, G., Domany, E., and Givol, D. Genome-wide comparison of human keratinocyte and squamous cell carcinoma responses to UVB irradiation: implications for skin and epithelial cancer, *Oncogene* in press, 2003.
11. Stupp, R., Dietrich, P.-Y., Ostermann Kraljevic, S., Pica, A., Maillard, I., Maeder, P., Meuli, R., Janzer, R., Pizzolato, G., Miralbell, R., Porchet, F., Regli, L., de Tribolet, N., Mirimanoff, R. O., and Leyvraz, S. Promising survival for patients with newly diagnosed glioblastoma multiforme treated with concomitant radiation plus temozolomide followed by adjuvant temozolomide, *J Clin Oncol*. 20: 1375-1382, 2002.
12. Ishii, N., Tada, M., Hamou, M. F., Janzer, R. C., Meagher-Villemure, K., Wiestler, O. D., Tribolet, N., and Van Meir, E. G. Cells with TP53 mutations in low grade astrocytic tumors evolve clonally to malignancy and are an unfavorable prognostic factor, *Oncogene*. 18: 5870-5878, 1999.

13. Albertoni, M., Shaw, P. H., Nozaki, M., Godard, S., Tenan, M., Hamou, M.-F., Fairlie, W. D., Breit, S. N., Paralkar, V. M., de Tribolet, N., Van Meir, E. G., and Hegi, M. E. Anoxia induces the macrophage inhibitory cytokine-1 (MIC-1) in glioblastoma cells independently of p53 and HIF-1, *Oncogene*. *21*: 4212-4219, 2002.
14. Flaman, J.-M., Frebourg, T., Moreau, V., Charbonnier, F., Martin, C., Chappuis, P., Sappino, A.-P., Limacher, J.-M., Bron, L., Benhattar, J., Tada, M., Van Meir, E. G., Estreicher, A., and Iggo, R. D. A simple p53 functional assay for screening cell lines, blood, and tumors., *Proc. Natl. Acad. Sci. USA*. *92*: 3963-3967, 1995.
15. Waridel, F., Estreicher, A., Bron, L., Flaman, J. M., Fontollet, C., Monnier, P., Frebourg, T., and Iggo, R. Field cancerisation and polyclonal p53 mutation in the upper aero-digestive tract., *Oncogene*. *14*: 163-169, 1997.
16. Nozaki, M., Tada, M., Kashiwazaki, H., Hamou, M.-F., Diserens, A.-C., Shinoe, Y., Sawamura, Y., Iwasaki, Y., de Tribolet, N., and Hegi, M. E. p73 is not mutated in meningiomas as determined with a functional yeast assay but p73 expression increases with tumor grade, *Brain Pathol*. *11*: 296-305, 2001.
17. Desbaillets, I., Diserens, A. C., Tribolet, N., Hamou, M. F., and Van Meir, E. G. Upregulation of interleukin 8 by oxygen-deprived cells in glioblastoma suggests a role in leukocyte activation, chemotaxis, and angiogenesis, *J Exp Med*. *186*: 1201-1212., 1997.
18. Plate, K. H., Breier, G., Weich, H. A., and Risau, W. Vascular endothelial growth factor is a potential tumour angiogenesis factor in human gliomas in vivo, *Nature*. *359*: 845-848, 1992.
19. Venables, W. N. and Ripley, B. D. *Modern Applied Statistics with S-PLUS*, 3rd edition edition, p. 333. New York: Springer Verlag, 1999.
20. Getz, G. and Domany, E. Coupled Two-Way Clustering Server, *Bioinformatics in press*, 2003.
21. Blatt, M., Wiseman, S., and Domany, E. Superparamagnetic clustering of data, *Phys Rev Lett*. *76*: 3251-3254, 1997.
22. Getz, G., Levine, E., Domany, E., and Zhang, M. Q. Superparamagnetic clustering of yeast gene expression profiles, *Physica A*. *279*: 457-464, 2000.
23. Kannan, K., Amariglio, N. R., G., Yaakov, Y., Kaminski, N., Getz, G., Domany, E., and Givol, D. Primary and secondary target genes regulated by p53 identified by DNA microarrays, *Oncogene*. *20*: 2225-2234, 2001.
24. Levine, E. and Domany, E. Resampling method for unsupervised estimation of cluster validity, *Neural Comput*. *13*: 2573-2593., 2001.
25. Benjamini, Y. and Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing, *J R Stat Soc B*. *57*: 289-300, 1995.
26. 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.
27. Ichimura, K., Bolin, M. B., Goike, H. M., Schmidt, E. E., Moshref, A., and Collins, V. P. Deregulation of the p14ARF/MDM2/p53 pathway is a prerequisite for human astrocytic gliomas with G1-S transition control gene abnormalities, *Cancer Res*. *60*: 417-424., 2000.

28. Grunstein, J., Roberts, W. G., Mathieu-Costello, O., Hanahan, D., and Johnson, R. S. Tumor-derived expression of vascular endothelial growth factor is a critical factor in tumor expansion and vascular function, *Cancer Res.* 59: 1592-1598., 1999.
29. Deuel, T. F., Zhang, N., Yeh, H. J., Silos-Santiago, I., and Wang, Z. Y. Pleiotrophin: a cytokine with diverse functions and a novel signaling pathway, *Arch Biochem Biophys.* 397: 162-171., 2002.
30. Choudhuri, R., Zhang, H. T., Donnini, S., Ziche, M., and Bicknell, R. An angiogenic role for the neurokines midkine and pleiotrophin in tumorigenesis, *Cancer Res.* 57: 1814-1819., 1997.
31. Feldser, D., Agani, F., Iyer, N. V., Pak, B., Ferreira, G., and Semenza, G. L. Reciprocal positive regulation of hypoxia-inducible factor 1alpha and insulin-like growth factor 2, *Cancer Res.* 59: 3915-3918., 1999.
32. Alfranca, A., Gutierrez, M. D., Vara, A., Aragonés, J., Vidal, F., and Landazuri, M. O. c-Jun and hypoxia-inducible factor 1 functionally cooperate in hypoxia-induced gene transcription, *Mol Cell Biol.* 22: 12-22., 2002.
33. Corps, A. N. and Brown, K. D. Insulin and insulin-like growth factor I stimulate expression of the primary response gene cMG1/TIS11b by a wortmannin-sensitive pathway in RIE-1 cells, *FEBS Lett.* 368: 160-164., 1995.
34. Bermont, L., Lamielle, F., Fauconnet, S., Esumi, H., Weisz, A., and Adessi, G. L. Regulation of vascular endothelial growth factor expression by insulin-like growth factor-I in endometrial adenocarcinoma cells, *Int J Cancer.* 85: 117-123., 2000.
35. Warren, R. S., Yuan, H., Matli, M. R., Ferrara, N., and Donner, D. B. Induction of vascular endothelial growth factor by insulin-like growth factor 1 in colorectal carcinoma, *J Biol Chem.* 271: 29483-29488., 1996.
36. Diviani, D. and Scott, J. D. AKAP signaling complexes at the cytoskeleton, *J Cell Sci.* 114: 1431-1437., 2001.
37. Grove, B. D. and Bruchey, A. K. Intracellular distribution of gravin, a PKA and PKC binding protein, in vascular endothelial cells, *J Vasc Res.* 38: 163-175., 2001.
38. O'Shea, J. J., Gadina, M., and Schreiber, R. D. Cytokine signaling in 2002: new surprises in the Jak/Stat pathway, *Cell.* 109: S121-131., 2002.
39. Dey, B. R., Furlanetto, R. W., and Nissley, P. Suppressor of cytokine signaling (SOCS)-3 protein interacts with the insulin-like growth factor-I receptor, *Biochem Biophys Res Commun.* 278: 38-43., 2000.
40. Deininger, M. H., Seid, K., Engel, S., Meyermann, R., and Schluesener, H. J. Allograft inflammatory factor-1 defines a distinct subset of infiltrating macrophages/microglial cells in rat and human gliomas, *Acta Neuropathol (Berl).* 100: 673-680., 2000.
41. Yin, Z. L., Dahlstrom, J. E., Le Couteur, D. G., and Board, P. G. Immunohistochemistry of omega class glutathione S-transferase in human tissues, *J Histochem Cytochem.* 49: 983-987., 2001.
42. Bodmer, S., Strommer, K., Frei, K., Siepl, C., de Tribolet, N., Heid, I., and Fontana, A. Immunosuppression and transforming growth factor-beta in glioblastoma. Preferential production of transforming growth factor-beta 2, *J Immunol.* 143: 3222-3229., 1989.

43. Walker, P. R. and Dietrich, P.-Y. Immune escape of gliomas. *In:* B. Castellano López and M. Nieto-Sampedro (eds.), *Progress in brain research*, Vol. 132, pp. 685-698: Elsevier Science B V, 2001.
44. Coons, S. W., Johnson, P. C., Scheithauer, B. W., Yates, A. J., and Pearl, D. K. Improving diagnostic accuracy and interobserver concordance in the classification and grading of primary glioma, *Cancer*. *79*: 1381-1393, 1997.
45. Kleihues, P. and Ohgaki, H. Primary and secondary glioblastomas: From concept to clinical diagnosis, *Neuro Oncol*. *1*: 44-51, 1999.
46. Elmlinger, M. W., Deininger, M. H., Schuett, B. S., Meyermann, R., Duffner, F., Grote, E. H., and Ranke, M. B. In vivo expression of insulin-like growth factor-binding protein-2 in human gliomas increases with the tumor grade, *Endocrinology*. *142*: 1652-1658., 2001.
47. Pintar, J. E., Schuller, A., Cerro, J. A., Czick, M., Grewal, A., and Green, B. Genetic ablation of IGFBP-2 suggests functional redundancy in the IGFBP family, *Prog Growth Factor Res*. *6*: 437-445., 1995.
48. Kerbel, R. and Folkman, J. Clinical translation of angiogenesis inhibitors, *Nat Rev Cancer*. *2*: 727-739., 2002.
49. Yancopoulos, G. D., Davis, S., Gale, N. W., Rudge, J. S., Wiegand, S. J., and Holash, J. Vascular-specific growth factors and blood vessel formation, *Nature*. *407*: 242-248., 2000.
50. Zadeh, G. and Guha, A. Neoangiogenesis in human astrocytomas: expression and functional role of angiopoietins and their cognate receptors, *Front Biosci*. *8*: E128-137., 2003.
51. Birner, P., Piribauer, M., Fischer, I., Gatterbauer, B., Marosi, C., Ambros, P. F., Ambros, I. M., Bredel, M., Oberhuber, G., Rössler, K., Budka, H., Harris, A. L., and Hainfellner, J. A. Vascular patterns in glioblastoma influence clinical outcome and associate with variable expression of angiogenic proteins: evidence for distinct angiogenic subtypes, *Brain Pathol*. *13*: 133-143, 2003.
52. Gruvberger, S., Ringner, M., Chen, Y., Panavally, S., Saal, L. H., Borg, A., Ferno, M., Peterson, C., and Meltzer, P. S. Estrogen receptor status in breast cancer is associated with remarkably distinct gene expression patterns, *Cancer Res*. *61*: 5979-584., 2001.
53. Welsh, J. B., Sapinoso, L. M., Su, A. I., Kern, S. G., Wang-Rodriguez, J., Moskaluk, C. A., Frierson, H. F., Jr., and Hampton, G. M. Analysis of gene expression identifies candidate markers and pharmacological targets in prostate cancer, *Cancer Res*. *61*: 5974-5978., 2001.
54. Fuller, G. N., Rhee, C. H., Hess, K. R., Caskey, L. S., Wang, R., Bruner, J. M., Yung, W. K., and Zhang, W. 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.
55. Sallinen, S. L., Sallinen, P. K., Haapasalo, H. K., Helin, H. J., Helen, P. T., Schraml, P., Kallioniemi, O. P., and Kononen, J. Identification of differentially expressed genes in human gliomas by DNA microarray and tissue chip techniques, *Cancer Res*. *60*: 6617-6622, 2000.
56. Rickman, D. S., Bobek, M. P., Misek, D. E., Kuick, R., Blaivas, M., Kurnit, D. M., Taylor, J., and Hanash, S. M. Distinctive molecular profiles of high-grade and low-grade gliomas based on oligonucleotide microarray analysis, *Cancer Res*. *61*: 6885-6891., 2001.

57. Hoeflich, A., Reisinger, R., Lahm, H., Kiess, W., Blum, W. F., Kolb, H. J., Weber, M. M., and Wolf, E. Insulin-like growth factor-binding protein 2 in tumorigenesis: protector or promoter?, *Cancer Res.* *61*: 8601-8610., 2001.
58. Bradshaw, S. L., D'Ercole, A. J., and Han, V. K. Overexpression of insulin-like growth factor-binding protein-2 in C6 glioma cells results in conditional alteration of cellular growth, *Endocrinology.* *140*: 575-584., 1999.
59. Fathallah-Shaykh, H. M., Rigen, M., Zhao, L. J., Bansal, K., He, B., Engelhard, H. H., Cerullo, L., Roenn, K. V., Byrne, R., Munoz, L., Rosseau, G. L., Glick, R., Lichtor, T., and DiSavino, E. Mathematical modeling of noise and discovery of genetic expression classes in gliomas, *Oncogene.* *21*: 7164-7174., 2002.

TABLES

Table 1 Verification of k-NN-class prediction using selected gene clusters on validation set of astrocytic gliomas

Comparisons	Classification 2-Way, k=3						Classification ^c 3-Way, k=3					
	ScGBM vs LGA		PrGBM vs ScGBM		PrGBM vs LGA		PrGBM vs (ScGBM+LGA)		LGA vs (ScGBM+PrGBM)		LGA vs PrGBM vs ScGBM	
	4 vs. 12		4 vs. 4		4 vs. 12		4 vs. (4+12)		12 vs. (4+4)		12 vs. 4 vs. 4	
Gene clusters used by the classifier ^a	No. of errors (%)	p-value ^b	No. of errors (%)	p-value	No. of errors (%)	p-value	No. of errors (%)	p-value	No. of errors (%)	p-value	Mean No of errors (%)	Mean p-value
G5	2 (12%)	0.050	1 (13%)	0.143	1 (6%)	0.0071	1 (5%)	0.00351	1 (5%)	0.00010	1-2 (8%)	5.70E-05
N505	2 (12%)	0.027	3 (37%)	1.000	2 (12%)	0.0269	4 (20%)	0.06089	3 (15%)	0.00322	6 (30%)	0.0022
N566	2 (12%)	0.050	2 (25%)	0.486	1 (6%)	0.0071	2 (10%)	0.01342	3 (15%)	0.00361	4 (20%)	0.0033
G12	3 (19%)	0.136	2 (25%)	0.486	2 (12%)	0.0269	3 (15%)	0.03199	7 (25%)	0.02489	5 (27%)	0.0484
G5+G505+G566+G12	1 (6%)	0.007	0 (0%)	0.029	0 (0%)	0.0005	0 (0%)	0.00021	0 (0%)	7.9E-06	1-2 (7%)	2.49E-06

^aThe classifier was built on the training set of 31 astrocytic gliomas, employing the respective gene set and the k-nearest neighbor (*k*-NN, *k*=3) method.

^bFisher exact test

^cClassification was repeated 100 times and p-values were averaged, see Materials and Methods for explanation.

FIGURES

Fig 1

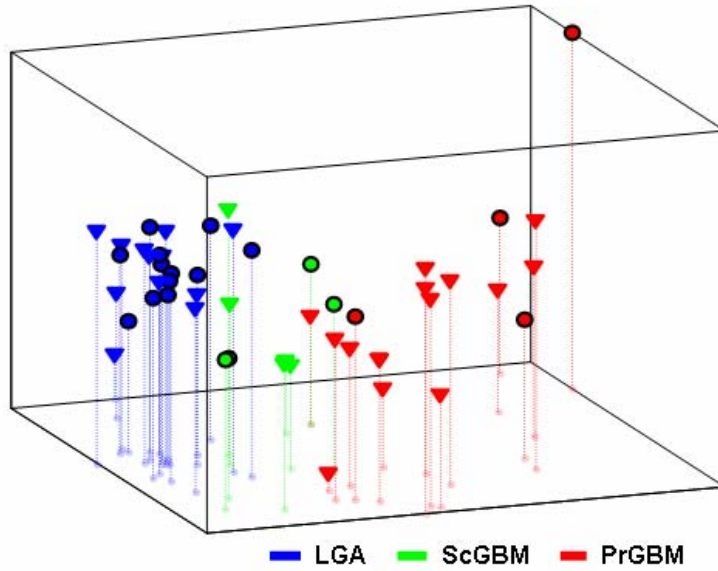


Fig. 1. Multidimensional scaling based on overall gene expression (1185 genes) of 51 astrocytic gliomas. The color code indicates the tumor subtype. Triangles represent samples from the training set, while circles denote gliomas from the validation set.

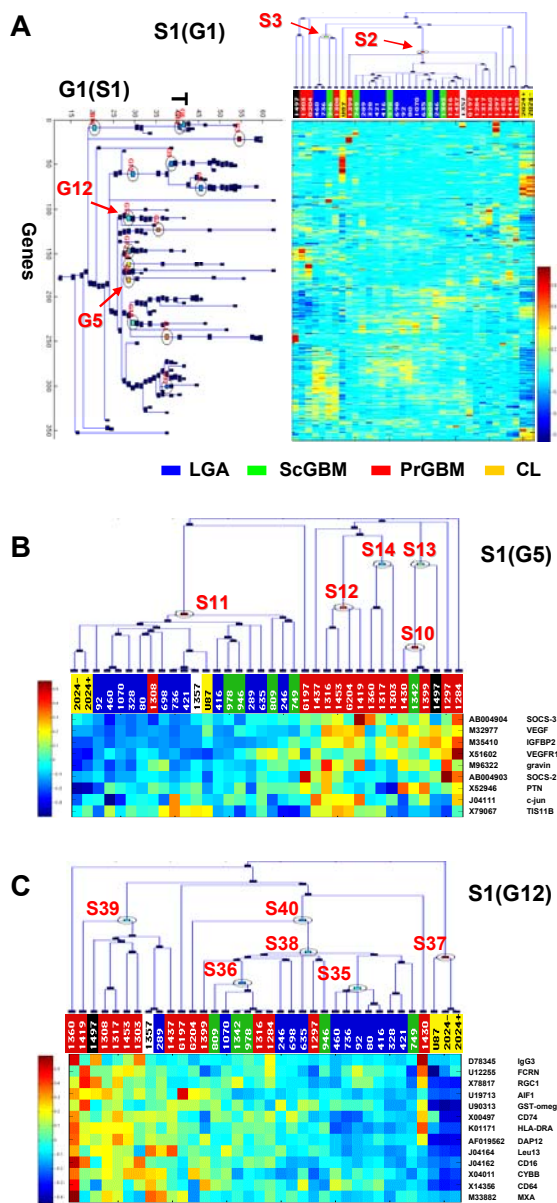


Fig 2

Fig. 2. Unsupervised analysis by Coupled Two-Way Clustering (CTWC). **(A) First level CTWC** clusters G1 (the set of all 358 genes that have passed the filtering criteria), using all 36 experiments (S1); and clusters all samples (S1) according to the gene expression profiles of all genes G1. These clustering operations are called G1(S1) and S1(G1), respectively. G1(S1) yields a dendrogram of gene clusters, and S1(G1) a dendrogram of sample clusters. The respectively reordered expression matrix is visualized using a color scale, representing centered and normalized values of the \log_2R . Fifteen stable gene clusters emerged and are marked with a ring (G2 to G16). Clustering the samples according to all genes, S1(G1), yielded two stable sample clusters, S2 and a smaller cluster S3. The tumor samples are denoted at the top, using colors to represent tumor subtypes: LGA, low grade astrocytoma; ScGBM, secondary glioblastoma; PrGBM, primary glioblastoma; RecGBM, recurrent glioblastoma; OAIH, oligoastrocytoma WHO grade III; CL, glioblastoma cell line. **(B) In Second level CTWC** we cluster all samples S1 according to selected gene clusters, S1(G5): Clustering S1 according to gene cluster G5 - (which was obtained from G1(S1) - see A) yields a dendrogram for the samples. Note the separation of LGA and ScGBM from PrGBM. PrGBM exhibit overexpression of the genes of G5 that are related to angiogenesis. **(C) Clustering S1** according to G12 yields a dendrogram S1(G12) separating a group of PrGBM in S39. G12 contains mostly genes related to the immune system. In the dendrograms a box represents a cluster. The T -value at which the box is placed corresponds to the temperature at which the cluster disintegrates. The sizes of the clusters are not reflected by their boxes. The parameters for a stable gene cluster have been set at a stability threshold of 8, a maximal drop out of 3 at a single step of T , and the minimal cluster size is 3. The criteria for stable sample clusters are: $Stab(C) > 8$, maximal drop out at a single step of T is 1, and minimal cluster size is 3.

Fig 3

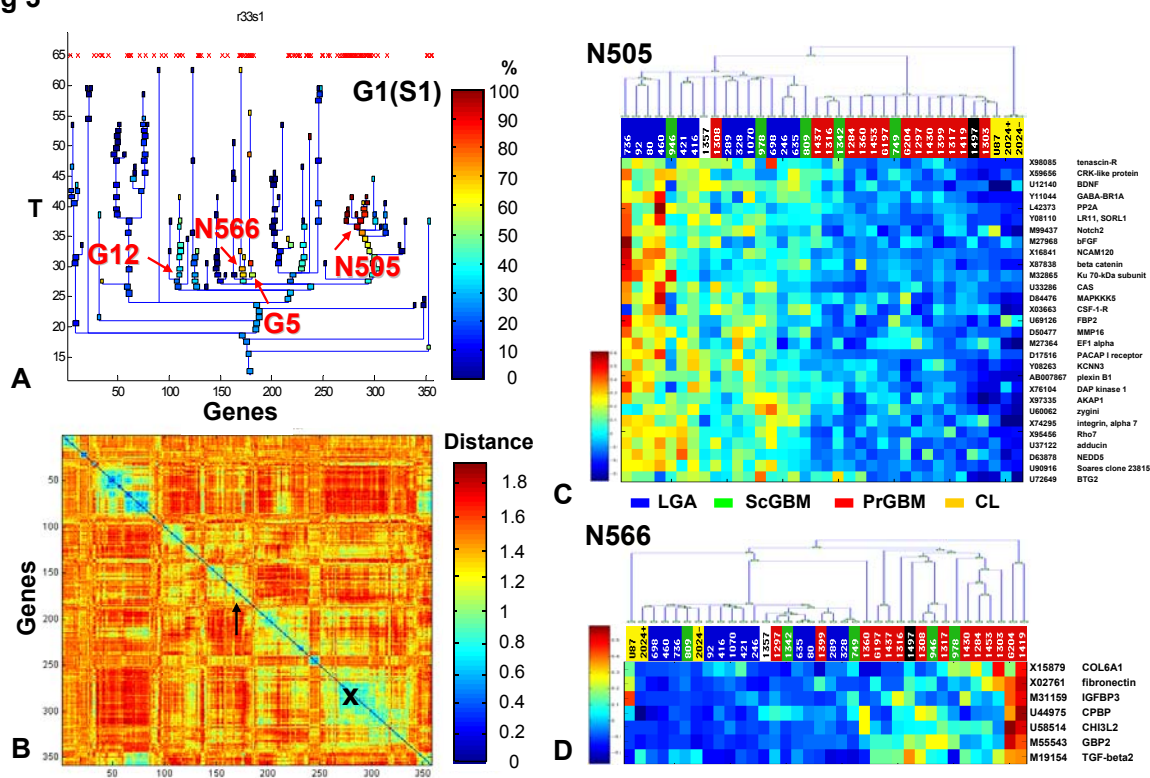


Fig. 3. Combination of supervised and unsupervised analysis. **(A)** The genes found by pairwise comparison of PrGBM vs LGA and ScGBM, were identified in the gene dendrogram G1(S1) (see also Figure 2A) and marked on top with a red cross. The percentage of these genes in a given gene cluster is presented in a color code. Most of the identified genes are grouped in 4 to 5 clusters. Two of them had been previously found, G5 and G12, marked with an arrow. The two new clusters identified are particularly rich in these genes, denoted as N505 and N566. They were not stable according to the criteria set, however, the distance matrix **(B)** recognizes these nodes as comprising closely related genes, indicated by an arrow for N566 and an X for N505. The distance D is visualized as a color code, with blue indicating short distance. Note, the order of the genes, and the scale is the same for **(A)** and **(B)**. Reclustering of all experiments S1 according to N505 and N566, respectively, is shown in **(C)** and **(D)**. 90% (27/29) of the N505 genes and >80% (6/7) of the N566 node were identified in the binary comparison, and therefore yield an almost perfect separation of the tumors according to their classification.

Fig 4

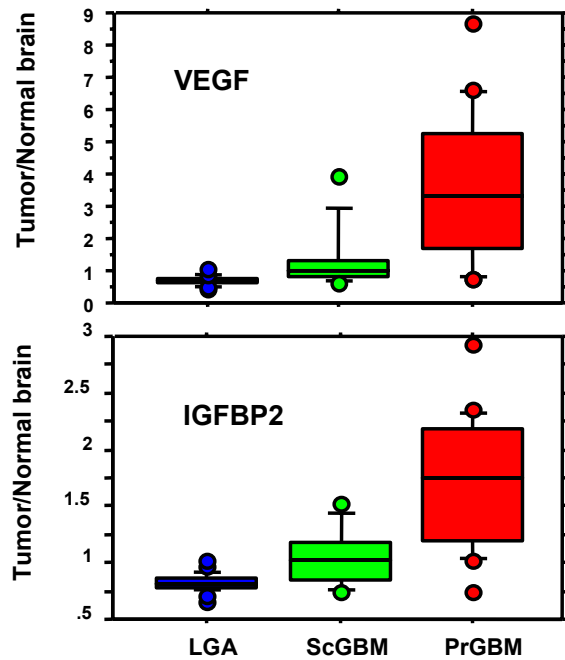


Fig. 4. Angiogenic activity correlates with tumor subtype. The relative expression of VEGF (**A**) and IGFBP2 (**B**) as compared to normal brain is shown for the 3 tumor subtypes, using all samples of the training and the validation set. Comparison between any two tumor subtypes showed highly significant differences for VEGF and IGFBP2, respectively (unpaired t-test: LGA vs.PrGBM, $p < 0.0001$ / $p < 0.0001$; LGA vs. ScGBM, $p = 0.0061$ / $p = 0.0003$; PrGBM vs. ScGBM, $p = 0.0099$ / $p = 0.0027$).

Fig 5

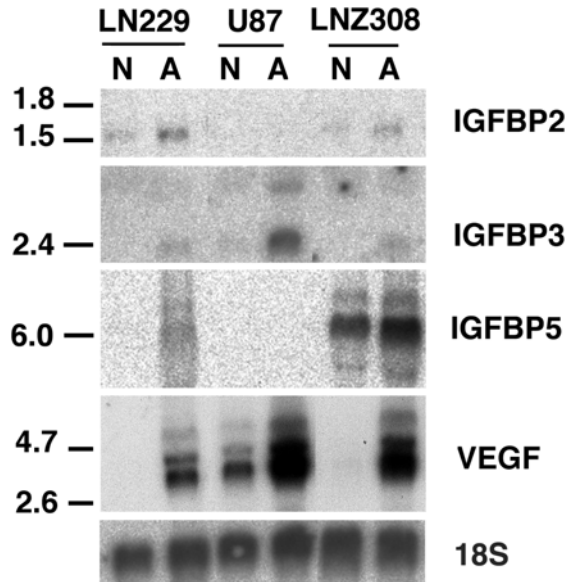


Fig. 5. Induction of IGFBP2, 3 or 5 by anoxia in glioblastoma cell lines. Glioblastoma cell lines LN229, U87, and LN-Z308 were cultured under normoxic (N) or anoxic (A) conditions for 20 hours. Expression of IGFBP2, 3 and 5 was evaluated by Northern blot analysis, and compared to induction of VEGF. At least one of the analyzed IGFBP family members was induced upon hypoxic treatment. Note, U87 the only cell line in this series harboring wild-type p53 exhibits substantial induction of the p53 target gene IGFBP3. Fragment sizes in kb are marked on the left.