

## Identification of diagnostic serum protein profiles of glioblastoma patients

Anja Elstner · Florian Stockhammer · Trong-Nghia Nguyen-Dobinsky · Quang Long Nguyen · Ingo Pilgermann · Amanjit Gill · Anke Guhr · Tingguo Zhang · Kajetan von Eckardstein · Thomas Picht · Julian Veelken · Robert L. Martuza · Andreas von Deimling · Andreas Kurtz

Received: 4 August 2008 / Accepted: 21 June 2010 / Published online: 9 July 2010  
© Springer Science+Business Media, LLC. 2010

**Abstract** Diagnosis of a glioblastoma (GBM) is triggered by the onset of symptoms and is based on cerebral imaging and histological examination. Serum-based biomarkers may support detection of GBM. Here, we explored serum protein concentrations of GBM patients and used data mining to explore profiles of biomarkers and determine whether these are associated with the clinical status of the patients. Gene and protein expression data for astrocytoma and GBM were used to identify secreted proteins differently expressed in tumors and in normal brain tissues. Tumor expression and serum concentrations of 14 candidate proteins were analyzed for 23 GBM patients and nine healthy subjects. Data-mining methods involving all 14 proteins were used as an

initial evaluation step to find clinically informative profiles. Data mining identified a serum protein profile formed by BMP2, HSP70, and CXCL10 that enabled correct assignment to the GBM group with specificity and sensitivity of 89 and 96%, respectively ( $p < 0.0001$ , Fischer's exact test). Survival for more than 15 months after tumor resection was associated with a profile formed by TSP1, HSP70, and IGFBP3, enabling correct assignment in all cases ( $p < 0.0001$ , Fischer's exact test). No correlation was found with tumor size or age of the patient. This study shows that robust serum profiles for GBM may be identified by data mining on the basis of a relatively small study cohort. Profiles of more than one biomarker enable more specific assignment to the GBM and survival group than those based on single proteins, confirming earlier attempts to correlate

A. Elstner and F. Stockhammer contributed equally to this work.

A. Elstner · A. Kurtz (✉)  
Berlin-Brandenburg Center for Regenerative Therapies, Charité-  
Universitätsmedizin Berlin, Augustenburger Platz 1, 13353  
Berlin, Germany  
e-mail: andreas.kurtz@charite.de

A. Elstner  
e-mail: anja.elstner@charite.de

F. Stockhammer · T. Picht  
Department of Neurosurgery, Charité-Universitätsmedizin  
Berlin, 13353 Berlin, Germany

F. Stockhammer  
e-mail: florian.stockhammer@charite.de

T. Picht  
e-mail: thomas.picht@charite.de

T.-N. Nguyen-Dobinsky · Q. L. Nguyen · A. Gill  
Medical Informatics and Bioinformatics, Charité-  
Universitätsmedizin Berlin, CCM, 10098 Berlin, Germany

T.-N. Nguyen-Dobinsky  
e-mail: trong-nghia.nguyen-dobinsky@charite.de

Q. L. Nguyen  
e-mail: nqlong@informatik.hu-berlin.de

A. Gill  
e-mail: amanjit.gill@gmx.de

I. Pilgermann  
Central Research Institute of Ambulatory Health Care,  
Herbert-Lewin-Platz 3, 10623 Berlin, Germany  
e-mail: pille.p@freenet.de

A. Guhr  
Licensing Office for hESC Research,  
Robert Koch Institute, 13353 Berlin, Germany  
e-mail: guhra@rki.de

T. Zhang · R. L. Martuza  
Department of Neurosurgery, Massachusetts General Hospital,  
Charlestown, MA 02129, USA

T. Zhang  
e-mail: ztguo@sdu.edu.cn

single markers with cancer. These conceptual findings will be a basis for validation in a larger sample size.

**Keywords** Data mining · Glioblastoma · Serum marker proteins

## Introduction

The glioblastoma (GBM) is a high-grade astrocytoma and is one of the most aggressive brain tumors. In most cases it is diagnosed only after clinical symptoms become apparent and depends on histological investigation of tumor samples obtained by biopsy or resection. Gene and protein expression profiling of astrocytic tumors, including cDNA microarray gene expression profiling of brain tumor tissue and cell lines [1–11] and quantitative proteomics analysis [12, 13], have recently been proposed as additional and refined diagnostic criteria. The analysis of microvesicles of glioma cells in culture revealed RNA and proteins that are potentially useful biomarkers [14]. Secreted proteins coded by astrocytoma-expressed genes may be detectable in the circulation and are thus candidates for noninvasive verification of diagnosis. Serum biomarkers have been detected by analysis of gene expression profiles of non-brain tumors, with promising results [15–17]. A few secreted proteins, including soluble CD95, YKL-40, frizzled, serum protein kinases, apolipoprotein E, cell adhesion molecules, and angiopoietins-1 and 2, have been proposed as potential serum biomarkers of astrocytomas [18–27]. None of these proteins alone was sufficiently specific and sensitive to serve as a diagnostic marker.

Because previous attempts to find surrogate serum markers for GBM failed when based on a single or only few

candidate factors, we attempted to analyze multiple factors to identify profiles associated with GBM. We screened astrocytoma gene and protein expression data bases for extracellular and secreted proteins differently expressed in astrocytoma. The serum concentrations of the identified proteins were determined for GBM patients by use of standard ELISA techniques. To identify a protein concentration pattern as a predictive model for the presence of a GBM, we chose to use unsupervised data mining as a first identification tool because it enables recognition of structures in the available data independent of a preset hypothesis [28]. Because of the limited sample size, we also used standard statistics for data analysis. The data show that serum profiles can be defined which are associated with GBM diagnosis and survival within the analyzed sample.

## Methods

The concentrations of 14 secreted candidate proteins were identified from data banks on astrocytoma gene expression. Tumor-specific protein expression was verified on tissue microarrays by immunohistochemistry. The concentrations of the protein were analyzed in the serum of 23 patients with GBM, and 12 control subjects. The data were profiled and associated with the disease by data mining and confirmed by bootstrapping. Filters were developed for serum protein profiles that enable diagnosis of GBM and prediction of survival of more or less than 15 months post GBM diagnosis within the study cohort.

### Patient recruitment

Twenty-three patients with GBM were recruited at the Departments of Neurosurgery at Massachusetts General Hospital, Boston, USA, Charité-Universitätsmedizin Berlin, Germany, Helios Klinikum Berlin-Buch, Germany, and Vivantes Klinikum Berlin-Neukölln, Germany. All tumor patients underwent gross tumor resection. Sera from nine healthy controls were acquired from the Department of Neurosurgery at Helios Klinikum Berlin. Control sera from three patients with rheumatoid arthritis treated with glucocorticoids (5–15 mg prednisolone/day) were obtained from the Department of Rheumatology and Clinical Immunology, Charité-Universitätsmedizin Berlin. The control group was age-adjusted to the study cohort (mean/median age for GBM group 55.3/58.6; mean/median age for control group 55.2/51.7). Clinical data and histopathological data for all patients were deposited in a phenotype-database (SecBase). Tumor volumes were approximated from MRI images using  $V(\text{mm}^3) = 4/3 \times \pi \times r(\text{sag}) \times r(\text{ax}) \times r(\text{cor})$ , or from removed tumor masses, where possible. Patient data are shown in Table 1. Peripheral blood was

R. L. Martuza  
e-mail: rmartuza@partners.org

K. von Eckardstein  
Department of Neurosurgery, Universitätsmedizin Göttingen,  
37099 Göttingen, Germany  
e-mail: kajetan.voneckardstein@med.uni-goettingen.de

J. Veelken  
Department of Neurosurgery, Vivantes Klinikum,  
12351 Berlin, Germany  
e-mail: julian.veelken@vivantes.de

A. von Deimling  
Department of Neuropathology, Ruprecht-Karls-Universität  
Heidelberg, and Clinical Cooperation Unit Neuropathology,  
German Cancer Research Center, 69120 Heidelberg, Germany  
e-mail: andreas.vondeimling@med.uni-heidelberg.de

A. Kurtz  
College of Veterinary Medicine, Seoul National University,  
Seoul, South Korea

**Table 1** Selected clinical data for GBM patients included in this study

No.	Survival (>15 months)	Tumor volume (cm <sup>3</sup> )	Survival (months)	Age	Radiation	Chemotherapy
10	Yes	32.3	34.6	55.0	Yes	Temozolomide
13	Yes	1.2	38.6	25.8	Yes	Temozolomide
12	Yes	n.d.	27.4	24.7	Yes	Adjuvant PCV
17	Yes	0.2	33.7	50.2	Yes	Temozolomide
18	Yes	2.4	43.6	78.8	Yes	Temozolomide
11	Yes	n.d.	44.7	36.2	yes	Carboplatin, thalidomide
19	Yes	4.2	58.9	53.7	Yes	Carboplatin, thalidomide
14	Yes	0.5	25.4	46.6	Yes	No
23	Yes	27	72.7	41.9	Yes	Temozolomide
5	Yes	54	21.8	66.7	Yes	Temozolomide
22	Yes	n.d.	20.0	63.9	Yes	No
7	Yes	n.d.	17.2	64.7	Yes	Temozolomide, PCV
	Mean	15.225	36.6	50.7		
15	No	n.d.	5.0	36.0	Yes	Temozolomide
16	No	2.4	2.5	69.8	Yes	Temozolomide
21	No	96	7.6	56.7	Yes	Temozolomide
20	No	1.4	5.6	64.5	Yes	Carmustine
8	No	50	7.1	68.6	Yes	No
3	No	1.2	14.2	42.6	Yes	Temozolomide
1	No	10	6.1	58.7	Yes	No
9	No	6.7	10.2	58.6	Yes	No
4	No	1.5	7.2	65.0	Yes	Temozolomide
2	No	n.d.	11.2	72.2	Yes	No
6	No	0.9	11.1	70.8	Yes	Temozolomide
	Mean	18.9	8.0	60.3		

All patients had diagnosed GBM IV. Diagnostic groups were divided according to survival less than or more than 15 months post surgery. All patients received radiation therapy  
PCV, combination of procarbazine, lomustine, and vincristine; PEV, combination of cisplatin, epirubicin, and vinorelbine; n.d., not determined

collected before surgery and serum was separated within 2–12 h, sub-divided, and stored at –80°C until use. We determined that no significant change in serum concentrations for the tested proteins occurred in this time window. All samples and data were acquired in accordance with applicable ethical standards and approved by the local ethics review boards.

Selection of candidate serum marker proteins

Genes potentially overexpressed in astrocytoma cells were identified by use of expression profiles in the serial analysis of gene expression (SAGE) database (<http://cgap.nci.nih.gov/SAGE>) using the Gene Expression Display tool as described [29]. The SAGE data for 24 astrocytoma and 12 normal brain samples were compared. RNA tags with more than a fivefold difference in their relative expression level in astrocytic tumors, and which are described by their Gene ontology (GO) annotations as coding for potentially secreted proteins, were selected. For analysis, a database was built in which the astrocytoma gene and protein expression levels were listed. GO terms were associated with each entry. In addition, clinical data, tumor features, and

previous reports on the presence of specific proteins in serum, where available, were recorded in the database. The GO terms used for selection of candidate secreted proteins were extracellular/extracellular space/matrix, cell surface receptor/integral to membrane, basement membrane, cell growth/maintenance, cell–cell-signaling/cell communication, cell motility, cell adhesion, cytokine activity, and the non-GO annotation serum factor. The terms were selected for their association with extracellular location and therefore enhanced likelihood of presence in the circulation [15, 16]. Candidate serum markers were further verified using published data on gene and protein overexpression in astrocytic tumors [1–5, 9, 12, 13, 19, 30–32].

Serum analysis

Serum samples were analyzed using commercially available ELISA kits for platelet factor 4 (PF4) (Diagnostica Stago, Asnieres-Sur-Seine, France), serotonin (IBL Immuno-Biological Laboratories, Hamburg, Germany), bone morphogenic protein 2 (BMP2), tumor necrosis factor-beta (TNF-β), stem cell factor (SCF/KITLG), insulin-like growth factor binding protein-3 (IGFBP3), chemokine (CXC motive)

ligand 10 (CXCL10), interleukin-1 $\alpha$  (IL-1 $\alpha$ ), fractalkine/chemokine (CX3C motive) ligand 1 (CX3CL1) (R&D Systems, Minneapolis, MN, USA), thrombospondin-1 (TSP1) (Chemicon International, Temecula, CA, USA), heat shock 70-kDa protein (HSP70) (Stressgen Biotechnologies, VIC, Canada), retinol-binding protein-4 (RBP4) (American Laboratory Products Company, Windham, USA), and midkine (MDK) (CellSignals, Yokohama, Japan). Serum concentrations for fatty acid binding protein-7 (FABP7) (Sanbio, Munich, Germany) were determined by slot-blot analysis (Bio-Rad Laboratories, Munich, Germany). All samples were tested at least in triplicate. Specific concentration standards for the tested factors were always included in the assays and standard adjustment of serum samples was performed before testing.

### Histology and immunohistochemistry

We compiled samples from two or three representative areas of 15 different GBM on a tissue micro array (TMA) as previously described [33]. Kidney, liver and non-neoplastic brain (temporal lobe tissue resected from epilepsy patients) specimens from different subjects served as controls. The TMA were probed with antibodies against HSP70 (Stressgen Biotechnologies), IGFBP3, TSP1 (Calbiochem, Bad Soden, Germany), BMP2 (R&D Systems), MK (CellSignals), and FABP-7 (kindly provided by Dr T. Müller, Berlin, Germany). Immunohistochemistry was performed using a BenchMark Automatic Staining Instrument (Ventana Medical Systems, Strasbourg, France). Briefly, paraffin sections were rehydrated and blocked in 10% serum. Automated cell conditioning for enhancement of antigen presentation was performed at 95°C for 8 min, at 100°C for 44 min, and at

42°C for 2 min. Primary antibodies were applied manually followed by incubation at 42°C for 30 min then subsequent automated application of biotinylated secondary antibody, horseradish peroxidase-labelled streptavidin, and DAB staining (Ventana Medical Systems) with intermediate washing steps. Finally, automated application of counterstain and bluing reagent (Ventana Medical Systems) was performed. Immunostaining was semiquantitatively scored by a histoscore index (percentage of positive cells  $\times$  staining intensity) in a blinded fashion.

### Data analysis

For detecting associations between the serum protein concentration data set and the clinical diagnosis and survival time data sets, the association rule discovery method [34] implemented by the software Magnum Opus of RuleQuest [35] was used. Magnum Opus employs unique *k-most-interesting* pattern-discovery techniques. We used the Windows version of Magnum Opus V2.3. Association rules found were selected manually to build decision trees for predictive engines [36]. In the first step all patient sets were combined to establish the decision tree (training set). The test sets are copies of the training set. Further, to validate the associations found by the applied method we performed bootstrapping, because this is generally superior to ANOVA for small data sets [37]. In this step we subsequently excluded one case from the training set and rebuilt the decision tree with the reduced training set. The excluded case was then used to test the reduced training set. The bootstrap results were obtained by repeating this procedure for all cases of the data set. They represented a well-validated and solid outcome.

**Table 2** Diagnostic candidate proteins selected for serum profiling in healthy and astrocytoma subjects

Candidate protein	Function	Mean serum concentration astrocytoma (SD)	Mean serum concentration control (SD)	<i>P</i> value ( <i>t</i> test)
HSP70 (ng/ml)	Anti-apoptotic	0.23 (39)	0.15 (0.06)	0.09
BMP2 (pg/ml)	Cell proliferation	209 (157)	176 (58.4)	0.33
RBP4 ( $\mu$ g/ml)	Retinoic-metabolism	64.2 (5.0)	39.4 (16.1)	0.02
IGFBP3 ( $\mu$ g/ml)	Pro-apoptotic	4.28 (1.5)	4.21 (1.2)	0.87
TSP1 ( $\mu$ g/ml)	Anti-angiogenic	77.7 (26.4)	67.2 (36.7)	0.37
CXCL10 (pg/ml)	Cell motility	34.2 (36)	105.4 (78.8)	0.001
CX3CL1 (ng/ml)	Cell motility	1.65 (1.9)	2.3 (3.4)	0.44
MDK (ng/ml)	Pro-angiogenic	739 (423)	243 (55)	0.07
TNFB (pg/ml)	Cytokine	0 (−46.7)	0 (0)	n.d.
SCF (pg/ml)	Cell proliferation/migration	440.6 (168.8)	728 (362.6)	0.032
IL1A (pg/ml)	Pro-inflammatory	0 (0.004)	0 (0.002)	n.d.
PF4 (UI/ml)	Anti-angiogenic	200 (33)	218.6 (28)	0.19
Serotonin (ng/ml)	Neurotransmitter	221.7 (85.4)	341.8 (88.4)	0.038
FABP7 (ng/ml)	Fatty acid binding	6.5	7.2	n.d.

Results are given as mean serum concentrations, the unit is indicated in the left row  
n.d., not determined

For statistical analysis of serum protein concentrations for each of the 14 single candidate proteins, a *t* test was applied. The validity of the test was calculated by use of Fisher’s exact test.

**Results**

Identification and selection of proteins potentially secreted by astrocytoma

SAGE expression data revealed 328 mRNA species highly expressed or underrepresented in astrocytomas compared with normal brain tissue. Thirty-six of these were identified as potentially astrocytoma-secreted transcripts based on GO-term assignments. Thirty-two proteins were identified by screening previously published gene and protein expression data from glioma [2–5, 9, 12, 13, 24, 30–32]. The final pool of candidate serum markers consisted of 68 proteins. Based on the availability of suitable detection systems 14 of the 68 candidate proteins were selected (Table 2).

Serum analysis of single candidate proteins

Analysis (*t* test) of serum protein concentrations for each of the 14 candidate proteins revealed raised serum concentrations in GBM patients compared with controls for HSP70 (*p* = 0.09), RBP4 (*p* = 0.02), serotonin (*p* = 0.04), and SCF (*p* = 0.03), and reduced serum concentrations for CXCL10 (*p* < 0.001). Slightly different serum levels were found for BMP2, TSP1, MDK, PF4, and CX3CL1 (*p* < 0.1). No differences were found between the two groups for IGFBP3, IL-1 $\alpha$ , and TNF- $\beta$ . For FABP7, results were not reproducible with the antibodies used. There was no correlation between the size of the tumor at diagnosis and the serum protein levels of any of the tested peptides, based on the limited number of subjects with available tumor volume (Table 1). No significant serum protein concentration difference was found between patients with rheumatoid arthritis (RA) and healthy subjects for the proteins tested (data not shown).

The GBM group was subdivided into two groups comprising patients surviving more than 15 month (*n* = 12) and those surviving less than 15 month post surgery (*n* = 11). Significant differences in serum protein concentration between these groups were found for TSP1 (*p* = 0.001) and IGFBP3 (*p* = 0.03) (*t* test).

Data mining analysis reveals protein profiles in GBM serum

Non-supervised data mining was used to propose potential diagnostic serum protein profiles consisting of at least

two proteins. Thresholds of serum protein concentrations for maximum differentiation between GBM and control group were: 208 pg/ml (BMP2), 0.24 ng/ml (HSP70), 3.8  $\mu$ g/ml (IGFBP3), 32.8  $\mu$ g/ml (TSP1), 33.9  $\mu$ g/ml

**Table 3** Potential diagnostic serum protein profile (BMP2, Hsp) associated with the clinical feature presence of a GBM

BMP2	Hsp70	CXCL10	Diagnosis	Prognosis	Results
737.8	160.799	7.6	GBM	GBM	True positive
303.1	63.709	42.0	GBM	GBM	True positive
632.5	9.853	23.4	GBM	GBM	True positive
123.8	5.077	16.9	GBM	GBM	True positive
223.8	37.714	29.8	GBM	GBM	True positive
393.1	1.943	139.7	GBM	GBM	True positive
227.1	<0.024	16.6	GBM	GBM	True positive
180.8	<0.024	19.0	GBM	GBM	True positive
83.8	0.351	34.7	GBM	GBM	True positive
109.1	<0.024	72.0	GBM	Healthy	False negative
225.1	0.950	53.6	GBM	GBM	True positive
185.8	0.072	7.4	GBM	GBM	True positive
257.1	0.044	24.12	GBM	GBM	True positive
107.1	0.649	146.9	GBM	GBM	True positive
117.9	0.351	6.8	GBM	GBM	True positive
114.6	<0.024	-0.17	GBM	GBM	True positive
373.8	0.064	11.4	GBM	GBM	True positive
158.8	0.344	19.9	GBM	GBM	True positive
161.3	0.538	81.8	GBM	GBM	True positive
117.1	0.138	19.3	GBM	GBM	True positive
107.1	0.520	13	GBM	GBM	True positive
132.9	0.210	38.6	GBM	GBM	True positive
112.1	1.270	21.5	GBM	GBM	True positive
207.8	<0.024	67.9	Healthy	Healthy	True negative
179.6	0.024	49.9	Healthy	Healthy	True negative
116.3	<0.024	102.8	Healthy	Healthy	True negative
125.4	0.061	93.7	Healthy	Healthy	True negative
140.8	<0.024	114.3	Healthy	Healthy	True negative
193.8	<0.024	65.5	Healthy	Healthy	True negative
132.1	0.052	76.6	Healthy	Healthy	True negative
120.4	0.217	153.4	Healthy	GBM	False positive
156.3	0.130	344.0	Healthy	Healthy	True negative

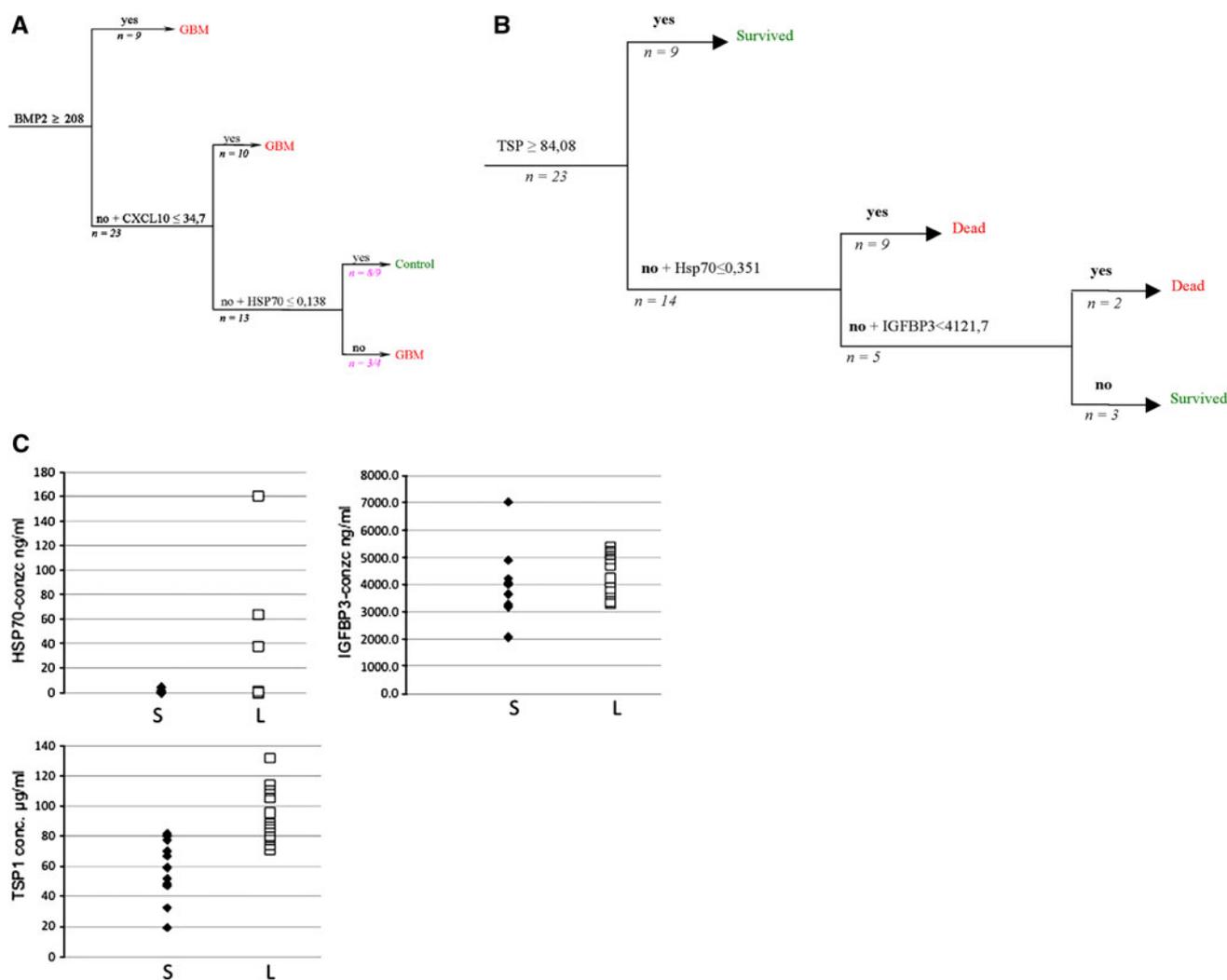
	Absolute	%
True positive	22	96
True negative	8	89
Sum true	30	94
False positive	1	11
False negative	1	4
Sum false	2	6

Serum concentrations were analyzed using commercially available ELISA kits. Protein profiles were identified by association analysis and confirmed by prognostic bootstrapping. Concentration for BMP-2 and CXCL10 in pg/ml. Concentration for HSP70 in ng/ml

(RBP4), 299 ng/ml (MDK), 1.1 ng/ml (CX3CL1), and 65.5 pg/ml (CXCL10). Except for CXCL1 and CXCL10, protein concentrations above these thresholds were associated with a GBM. No thresholds were found for the concentrations of the remaining proteins (serotonin, SCF, MDK, FABP7, PF4, IL-1 $\alpha$ , and TNF- $\beta$ ).

The defined threshold concentrations were subsequently used for identification of protein profiles by association analysis. The serum profile formed by BMP2, CXCL10, and HSP70 was associated with the clinical feature presence of GBM (Table 3, Fig. 1a). The profile correctly assigned 96% of the GBM subjects and 89% of control subjects by bootstrap validation ( $p < 0.0001$ , Fisher's exact test).

For the clinical feature 15 months survival post surgery, association was achieved with serum profiles based on concentrations of TSP1, HSP70, and IGFBP3. Whereas analysis of TSP1 alone predicts the survival chance for 80% of the GBM patients, inclusion of the other two factors increased predictability to 100%, as validated by a bootstrap algorithm ( $p < 0.0001$ , Fisher's exact test) (Table 4, Fig. 1b, c). The two survival groups were not distinguished by different treatment schemes, because these were comparable—standard palliative therapy including radio and chemotherapy was used for both groups (Table 1). There was also no association between tumor size and survival. Although the mean age for the long-term survivor group was 50.7 years compared with 60.3 years for the short-term



**Fig. 1** Decision trees. **a** Clinical feature presence of a GBM. Potential diagnostic serum protein profile BMP2, CXCL10, and HSP70. Concentrations of BMP-2 and CXCL10 in pg/ml. Concentration of HSP70 in ng/ml. **b** Clinical feature 15 months survival post surgery. Potential diagnostic serum protein profile TSP1, HSP70, and

IGFBP3. Concentrations of TSP1 and IGFBP3 in µg/ml. Concentration of Hsp70 in ng/ml. **c** Comparison of serum concentrations for Hsp70, IGFBP3, and TSP1 between the groups S (survival <15 months) and L (survival more than 15 months) of the study cohort

**Table 4** Potential diagnostic serum protein profile (TSP1, HSP70, and IGFBP3) associated with the clinical feature 15-month survival post surgery

Hsp70	IGFBP3	TSP	Diagnosis	Prognosis	Results
160.799	3891.4	88.34	Survived	Survived	True negative
63.709	3753.8	110.1	Survived	Survived	True negative
9.853	4273.7	88.24	Survived	Survived	True negative
5.077	3280.7	70.3	Dead	Dead	True positive
37.714	5130.7	74.3	Survived	Survived	True negative
1.943	2080.4	47.56	Dead	Dead	True positive
<0.024	3175.8	67.2	Dead	Dead	True positive
<0.024	3208.4	59.6	Dead	Dead	True positive
0.351	4075.5	52.3	Dead	Dead	True positive
<0.024	4229.4	32.9	Dead	Dead	True positive
0.950	3327.3	113.93	Survived	Survived	True negative
0.072	4019.6	77.61	Dead	Dead	True positive
0.044	4907.7	19.84	Dead	Dead	True positive
0.649	5229.4	79.2	Survived	Survived	True negative
0.351	4963.7	95.6	Survived	Survived	True negative
<0.024	3658.3	80.3	Dead	Dead	True positive
0.064	4264.3	131.99	Survived	Survived	True negative
0.344	3766.5	105.16	Survived	Survived	True negative
0.538	1695.7	84.08	Survived	Survived	True negative
0.138	7039.7	48.4	Dead	Dead	True positive
0.520	3411.8	86.08	Survived	Survived	True negative
0.210	3273.3	82.12	Dead	Dead	True positive
1.270	5398.6	71.14	Survived	Survived	True negative
			Absolute		%
True positive			11		100
True negative			12		100
Sum true			23		100
False positive			0		0
False negative			0		0
Sum false			0		0

Serum concentrations were analyzed using commercially available ELISA kits. Protein profiles were identified by association analysis and confirmed by prognostic bootstrapping. Concentrations of TSP1 and IGFBP3 in µg/ml. Concentration of HSP70 in ng/ml

group (non-significant,  $p = 0.095$ ,  $t$  test), the serum protein profile did not correlate with age.

**Immunohistochemical detection of secreted candidate proteins in glioblastoma**

Immunohistochemistry showed strong cytosolic expression of HSP70 and FABP7 in most GBM (Fig. 2). Some tumor cells showed nuclear presence of FABP7. Strong nuclear and perinuclear immunostaining of tumor cells was detected for TSP1, whereas IGFBP3 was expressed moderately in the cytoplasm. MDK expression was diffusely

present in a minority of the tumors. BMP2 expression was negligible (not shown). In control brain sections IGFBP3 was not detectable. HSP70 was only detected in neurons and in a small number of astrocytes, TSP1 was strongly expressed and FABP7 expression was found in reactive astrocytes. Histoscore data for individual tumors and serum levels of matched samples did not correlate, even when tumor size was taken into account (not shown).

**Discussion**

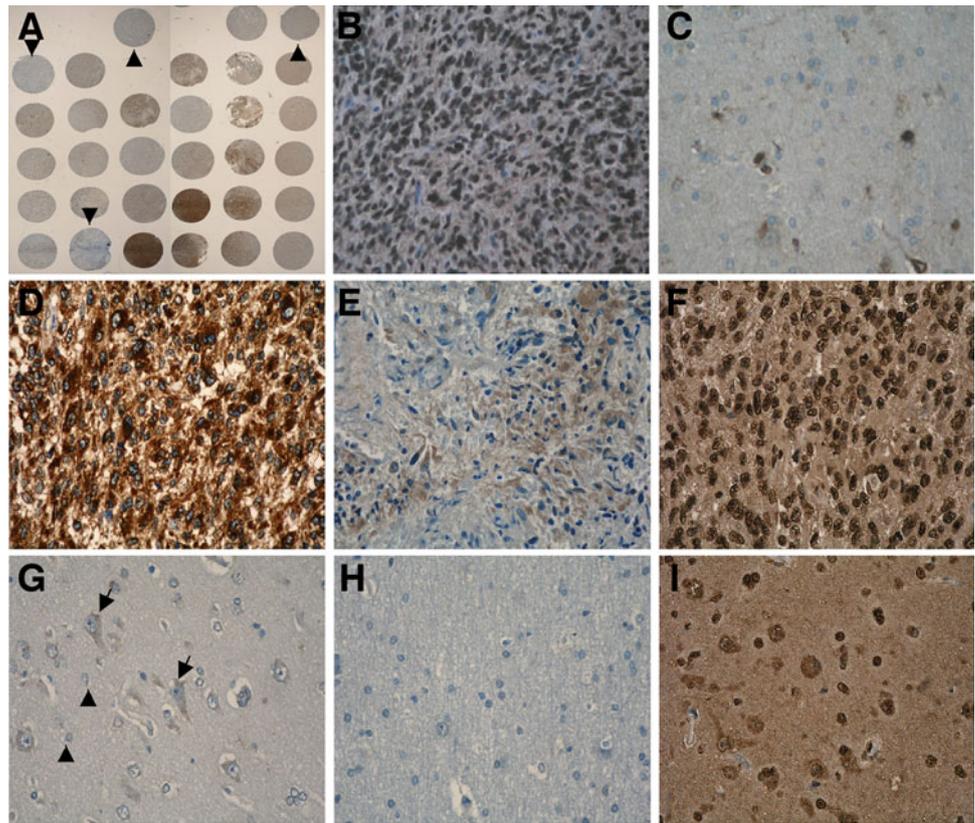
This is the first study to search systematically for a diagnostic serum profile in GBM patients. Single molecular markers are mostly insufficient to follow the dynamics of diseases such as cancer, and will be replaced by multiple marker profiles [38]. The number of proteins needed for a defined profile is not known, and most likely varies between diseases. In gene expression-based diagnostics, profiles often consist of <10, but some times of more than 20 genes [32, 39]. To obtain an estimate of this number in GBM, we analyzed serum samples of a small cohort of GBM patients and controls and adopted data mining to detect protein profiles. The data mining method used here delivered a testable hypothesis.

Currently, few secreted proteins expressed by astrocytomas have been proposed as potential serum markers [18–27]. When tested in serum, none of these was sufficiently specific to serve as a diagnostic marker. We evaluated the serum concentrations of 14 proteins, which were selected by screening gene and protein expression profiles of astrocytomas for proteins potentially released by the tumor cells [15]. Included were also two classes of cytoplasmic proteins associated with cell stress (HSP70) [40] and with neural stem cells (FABP7) [41]. Members of both protein classes have previously been found in serum and proposed as biomarkers [42, 43]. More generally, evaluation of the serum proteome revealed that numerous typically cytoplasmic proteins are found in serum [44].

In contrast with single biomarkers, in a complex diagnostic profile of multiple proteins it may not be necessary for the concentrations of each protein to be significantly different between two groups. Our results are in agreement with this assumption. Using the complete dataset to identify patterns of proteins by applying data mining, it was possible to associate combinations of serum proteins with the clinical diagnosis “presence of a GBM”. A profile with a relatively small number of proteins (BMP2, CXCL10, HSP70) was sufficient to correctly assign 96% of the GBM and 89% of the control subjects. One subject from the GBM and one from the control group were misdiagnosed using this profile.

The proteins constituting the identified GBM profiles are functionally diverse and not directly correlated with

**Fig. 2** Tissue array based detection of protein expression in astrocytoma by immunohistochemistry. Magnification is  $\times 400$  for all sections. **a** Overview section of the tissue array used after detection of HSP70 by immunostaining. Each section represents an individual tumor or control tissue. Non-tumor control tissues are labeled with *arrowheads*. **b, c** FABP7 detection in **b** GBM and **c** control brain tissues. **d, g** HSP70 detection in **d** GBM and **g** control brain tissues. *Arrows* neurons, *arrowheads* glial cells. **e, h** IGFBP3 detection in **e** GBM and **h** control brain tissues. **f, i** TSP1 detection in **f** GBM and **i** control brain tissues



tumor status. The interferon- $\gamma$  (IFN- $\gamma$ ) inducible chemokine CXCL10, concentrations of which were reduced in GBM patients, attracts CXCR3-receptor-carrying cells. Besides its possible transmitter function in the brain [45], it inhibits angiogenesis and has anti-tumor activity in vivo [46] and its reduced concentration may thus provide a growth advantage for GBM. HSP70 is an anti-apoptotic chaperone with tumor-promoting activity and has been described as overexpressed in GBM, although its overall role in this type of cancer remains unclear [47]. Increased HSP70 antibody levels have been found in serum of lung cancer patients, indicating that it may be a biomarker for several tumor types [48]. BMP2 is a member of the transforming growth factor beta (TGFbeta) family, which inhibits growth in gliomas [49] and medulloblastomas [50], whereas the latter tumors are able to secrete BMP2 themselves. On the other hand, BMP2 is linked to angiogenesis [51, 52], which is consistent with the glioblastoma being one of the most vascularized tumors in man [53]. The role of BMP2 in glioblastoma has not yet been elucidated. TSP1 has tumor-inhibitory properties because of its anti-angiogenic function. Reduced serum levels of TSP1 have been reported in small-cell lung cancer patients [54]. Upregulation of TSP1 has been shown to be expressed in microvascular hyperplasia during astrocytoma progression [55] and may explain the increased serum levels found in

patients with short survival. IGFBP3 mediates p53-dependent apoptosis in GBM [56] and was associated with enhanced survival in the GBM group in this study.

Our data suggest a method for hypothesis generation and provide only a first indication of the presence of GBM specific serum profiles. Prospective and retrospective validation of the data on larger sample sizes will be necessary to evaluate and, perhaps, adjust the presented profiles further.

## Conclusion

The study shows that robust serum profiles for GBM are identifiable by data mining based on a relatively small study cohort. The presence of a glioblastoma was distinguished by serum concentrations of BMP2, CXCL10, and HSP70 with a sensitivity of 96% and specificity of 89% using the decision tree. The TSP1, HSP70, and IGFBP3 serum profile could be used to assign a survival prognosis in our data set. These findings will be a basis for validation with a larger sample size and with recurrent glioma.

**Acknowledgments** We wish to thank Dr Lydia Naumann (Charité, Berlin, Germany) for providing serum samples, Dr Sadatoshi Sakuma (CellSignals, Japan) for kindly providing MK antibodies and ELISA kit, and Dr Thomas Müller (MDC Berlin-Buch, Germany) for providing FABP7 antibodies. We are grateful to Drs David Louis,

Cathrin Nutt, and Samuel Rabkin (MGH, Boston, USA) for their continuous support of this study and their valuable suggestions. A.E. and F.S. were supported by the Berliner Krebsgesellschaft e.V. The study was funded by the Goldhirsh Foundation, Boston, USA.

## References

- Sallinen SL, Sallinen PK, Haapasalo HK et al (2000) Identification of differentially expressed genes in human gliomas by DNA microarray and tissue chip techniques. *Cancer Res* 60(23):6617–6622
- Rickman DS, Bobek MP, Misek DE et al (2001) Distinctive molecular profiles of high-grade and low-grade gliomas based on oligonucleotide microarray analysis. *Cancer Res* 61(18):6885–6891
- Fuller GN, Hess KR, Rhee CH et al (2002) Molecular classification of human diffuse gliomas by multidimensional scaling analysis of gene expression profiles parallels morphology-based classification, correlates with survival, and reveals clinically-relevant novel glioma subsets. *Brain Pathol* 12(1):108–116
- Nutt CL, Mani DR, Betensky RA et al (2003) Gene expression-based classification of malignant gliomas correlates better with survival than histological classification. *Cancer Res* 63(7):1602–1607
- Shai R, Shi T, Kremen TJ et al (2003) Gene expression profiling identifies molecular subtypes of gliomas. *Oncogene* 22(31):4918–4923
- van den Boom J, Wolter M, Kuick R et al (2003) 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(3):1033–1043
- Liang Y, Diehn M, Watson N et al (2005) Gene expression profiling reveals molecularly and clinically distinct subtypes of glioblastoma multiforme. *Proc Natl Acad Sci USA* 102(16):5814–5819
- Hess KR, Fuller GN, Rhee CH et al (2001) Statistical pattern analysis of gene expression profiles for glioblastoma tissues and cell lines. *Int J Mol Med* 8(2):183–188
- Watson MA, Perry A, Budhraj V et al (2001) Gene expression profiling with oligonucleotide microarrays distinguishes World Health Organization grade of oligodendrogliomas. *Cancer Res* 61(5):1825–1829
- Reddy SP, Britto R, Vinnakota K et al (2008) Novel glioblastoma markers with diagnostic and prognostic value identified through transcriptome analysis. *Clin Cancer Res* 14(10):2978–2987
- Shirahata M, Iwao-Koizumi K, Saito S et al (2007) Gene expression-based molecular diagnostic system for malignant gliomas is superior to histological diagnosis. *Clin Cancer Res* 13(24):7341–7356
- Furuta M, Weil RJ, Vortmeyer AO et al (2004) Protein patterns and proteins that identify subtypes of glioblastoma multiforme. *Oncogene* 23(40):6806–6814
- Iwadate Y, Sakaida T, Hiwasa T et al (2004) Molecular classification and survival prediction in human gliomas based on proteome analysis. *Cancer Res* 64(7):2496–2501
- Skog J, Würdinger T, van Rijn S, Meijer DH, Gainche L, Sena-Esteves M, Curry WT Jr, Carter BS, Krichevsky AM, Breakefield XO (2008) Glioblastoma microvesicles transport RNA and proteins that promote tumour growth and provide diagnostic biomarkers. *Nat Cell Biol* 10(12):1470–1476
- Welsh JB, Sapinoso LM, Kern SG et al (2003) Large-scale delineation of secreted protein biomarkers overexpressed in cancer tissue and serum. *Proc Natl Acad Sci USA* 100(6):3410–3415
- Diehn M, Eisen MB, Botstein D et al (2000) Large-scale identification of secreted and membrane-associated gene products using DNA microarrays. *Nat Genet* 25(1):58–62
- Zhou W, Sokoll LJ, Bruzek DJ et al (1998) Identifying markers for pancreatic cancer by gene expression analysis. *Cancer Epidemiol Biomarkers Prev* 7(2):109–112
- Audero E, Cascone I, Zanon I et al (2001) Expression of angiopoietin-1 in human glioblastomas regulates tumor-induced angiogenesis: in vivo and in vitro studies. *Arterioscler Thromb Vasc Biol* 21(4):536–541
- Tanwar MK, Gilbert MR, Holland EC (2002) Gene expression microarray analysis reveals YKL-40 to be a potential serum marker for malignant character in human glioma. *Cancer Res* 62(15):4364–4368
- Streffer JR, Schuster M, Zipp F et al (1998) Soluble CD95 (Fas/APO-1) in malignant glioma: (no) implications for CD95-based immunotherapy? *J Neurooncol* 40(3):233–235
- Krul ES, Tang J (1992) Secretion of apolipoprotein E by an astrocytoma cell line. *J Neurosci Res* 32(2):227–238
- Koshikawa N, Nakamura T, Tsuchiya N et al (1996) Purification and identification of a novel and four known serine proteinase inhibitors secreted by human glioblastoma cells. *J Biochem* 119(2):334–339
- Roth W, Wild-Bode C, Platten M et al (2000) Secreted Frizzled-related proteins inhibit motility and promote growth of human malignant glioma cells. *Oncogene* 19(37):4210–4220
- Mishima K, Asai A, Kadomatsu K et al (1997) Increased expression of midkine during the progression of human astrocytomas. *Neurosci Lett* 233(1):29–32
- Nano R, Capelli E, Argentina F et al (2003) Evaluation of serum levels of cytokines and intercellular adhesion molecule-1 (ICAM-1) in astrocytic tumours. *Cell Mol Biol (Noisy-le-grand)* 49(4):525–528
- Salmaggi A, Eoli M, Frigerio S et al (1999) Circulating intercellular adhesion molecule-1 (ICAM-1), vascular cell adhesion molecule-1 (VCAM-1) and plasma thrombomodulin levels in glioblastoma patients. *Cancer Lett* 146(2):169–172
- Koga K, Todaka T, Morioka M et al (2001) Expression of angiopoietin-2 in human glioma cells and its role for angiogenesis. *Cancer Res* 61(16):6248–6254
- Fayyad JA, Jahshan CS, Karam EG (2001) Systems development of child mental health services in developing countries. *Child Adolesc Psychiatr Clin N Am* 10(4):745–762 (ix)
- Kavsan V, Shostak K, Dmitrenko V et al (2005) Characterization of genes with increased expression in human glioblastomas. *Tsitol Genet* 39(6):37–49
- Freije WA, Castro-Vargas FE, Fang Z et al (2004) Gene expression profiling of gliomas strongly predicts survival. *Cancer Res* 64(18):6503–6510
- Ehteshami M, Yuan X, Kabos P et al (2004) Glioma tropic neural stem cells consist of astrocytic precursors and their migratory capacity is mediated by CXCR4. *Neoplasia* 6(3):287–293
- Kim JH, Skates SJ, Uede T et al (2002) Osteopontin as a potential diagnostic biomarker for ovarian cancer. *JAMA* 287(13):1671–1679
- Kononen J, Bubendorf L, Kallioniemi A et al (1998) Tissue microarrays for high-throughput molecular profiling of tumor specimens. *Nat Med* 4(7):844–847
- Webb G, Zhang S (2002) Beyond association rules: generalized rule discovery. Kluwer, Dordrecht
- Webb G (1995) OPUS: an efficient admissible algorithm for unordered search. *J Artif Intell Res* 3:431–465
- Doddi S, Marathe A, Ravi SS et al (2001) Discovery of association rules in medical data. *Med Inform Internet Med* 26(1):25–33
- Davidson AC, Hinkley DV (1997) Bootstrap methods and their application. Cambridge University Press, Cambridge

38. Johann DJ Jr, McGuigan MD, Patel AR et al (2004) Clinical proteomics and biomarker discovery. *Ann N Y Acad Sci* 1022: 295–305
39. Bucca G, Carruba G, Saetta A et al (2004) Gene expression profiling of human cancers. *Ann N Y Acad Sci* 1028:28–37
40. Naylor DJ, Hartl FU (2001) Contribution of molecular chaperones to protein folding in the cytoplasm of prokaryotic and eukaryotic cells. *Biochem Soc Symp* 68:45–68
41. Kurtz A, Zimmer A, Schnutgen F et al (1994) The expression pattern of a novel gene encoding brain-fatty acid binding protein correlates with neuronal and glial cell development. *Development* 120(9):2637–2649
42. Pockley AG, Georgiades A, Thulin T et al (2003) Serum heat shock protein 70 levels predict the development of atherosclerosis in subjects with established hypertension. *Hypertension* 42(3): 235–238
43. Chan CP, Sanderson JE, Glatz JF et al (2004) A superior early myocardial infarction marker. Human heart-type fatty acid-binding protein. *Z Kardiol* 93(5):388–397
44. Chan KC, Lucas DA, Hise D et al (2004) Analysis of the human serum proteome. *Clin Proteom* 1(2):26
45. Adler MW, Rogers TJ (2005) Are chemokines the third major system in the brain? *J Leukoc Biol* 78(6):1204–1209
46. Angiolillo AL, Sgadari C, Taub DD et al (1995) Human interferon-inducible protein 10 is a potent inhibitor of angiogenesis in vivo. *J Exp Med* 182(1):155–162
47. Strik HM, Weller M, Frank B et al (2000) Heat shock protein expression in human gliomas. *Anticancer Res* 20(6B):4457–4462
48. Zhong L, Peng X, Hidalgo GE et al (2003) Antibodies to HSP70 and HSP90 in serum in non-small cell lung cancer patients. *Cancer Detect Prev* 27(4):285–290
49. Chattopadhyay N, TFH J, Godbole MM et al (2004) Transforming growth factor beta receptor family ligands inhibit hepatocyte growth factor synthesis and secretion from astrocytoma cells. *Brain Res Mol Brain Res* 121(1–2):146–150
50. Hallahan AR, Pritchard JI, Chandraratna RA et al (2003) BMP-2 mediates retinoid-induced apoptosis in medulloblastoma cells through a paracrine effect. *Nat Med* 9(8):1033–1038
51. Yang S, Pham LK, Liao CP et al (2008) A novel bone morphogenetic protein signaling in heterotypic cell interactions in prostate cancer. *Cancer Res* 68(1):198–205
52. Landry Y, Gies JP (2008) Drugs and their molecular targets: an updated overview. *Fundam Clin Pharmacol* 22(1):1–18
53. Brem S, Tsanaclis AM, Gately S et al (1992) Immunolocalization of basic fibroblast growth factor to the microvasculature of human brain tumors. *Cancer* 70(11):2673–2680
54. Dudek AZ, Mahaseth H (2005) Circulating angiogenic cytokines in patients with advanced non-small cell lung cancer: correlation with treatment response and survival. *Cancer Invest* 23(3):193–200
55. Brat DJ, Kaur B, Van Meir EG (2003) Genetic modulation of hypoxia induced gene expression and angiogenesis: relevance to brain tumors. *Front Biosci* 8:d100–d116
56. Shen L, Dean NM, Glazer RI (1999) Induction of p53-dependent, insulin-like growth factor-binding protein-3-mediated apoptosis in glioblastoma multiforme cells by a protein kinase Calpha antisense oligonucleotide. *Mol Pharmacol* 55(2):396–402