Identifying Potential Tumor Markers and Antigens by Database Mining and Rapid Expression Screening

W. Troy Loging,[1,5] Anita Lal,[1,5] I-Mei Siu,[1] Tania L. Loney,[1] Carol J. Wikstrand,[1] Marco A. Marra,[2] Christa Prange,[3] Darell D. Bigner,[1] Robert L. Strausberg,[4] and Gregory J. Riggins[1,6]

[1]Duke University Medical Center, Durham, North Carolina 27710, USA; [2]Washington University Genome Sequencing Center, St. Louis, Missouri 63108, USA; [3]The I.M.A.G.E. Consortium, Biology and Biotechnology Research Program, Lawrence Livermore National Laboratory, Livermore, California 94550, USA; [4]Cancer Genome Anatomy Project, Office of the Director, National Cancer Institute, Bethesda, Maryland 20892, USA

Genes expressed specifically in malignant tissue may have potential as therapeutic targets but have been difficult to locate for most cancers. The information hidden within certain public databases can reveal RNA transcripts specifically expressed in transformed tissue. To be useful, database information must be verified and a more complete pattern of tissue expression must be demonstrated. We tested database mining plus rapid screening by fluorescent-PCR expression comparison (F-PEC) as an approach to locate candidate brain tumor antigens. Cancer Genome Anatomy Project (CGAP) data was mined for genes highly expressed in glioblastoma multiforme. From 13 mined genes, seven showed potential as possible tumor markers or antigens as determined by further expression profiling. Now that large-scale expression information is readily available for many of the commonly occurring cancers, other candidate tumor markers or antigens could be located and evaluated with this approach.

[The expression data described in this paper have been submitted to the NCBI SAGEmap database under library name SAGE_Duke_GBM_H1110, SAGE_pooled_GBM, SAGE_BB542_whitematter, and SAGE_normal_pool(6th).]

During malignant progression, the pattern of expressed genes can provide clues to understanding tumor growth. In addition to insight into the tumor biology that might be derived from this pattern, there is a practical application for identifying genes highly expressed in tumors but not in normal adult tissue. A common example of tumor marker use is the serum protein assay for early detection of cancer (Kardamakis 1996). Investigators are also searching for genomic DNA alterations or abnormal gene expression in other clinically accessible samples. Progress has been made on finding tumor markers in stool (Sidransky et al. 1992; Vogelstein and Kinzler 1999), sputum (Mao et al. 1994), and urine (Lokeshwar et al. 1997).

Tumor-specific gene expression may also provide an opportunity for immune-based cancer therapies by targeting one or more of the tumor antigens coded for by these genes. Toxic antibodies with high affinity to accessible cell surface or extracellular proteins may kill enough cancer cells to be therapeutic (Panchal 1998). Recent success with monoclonal antibody targeting of

the Her/neu-2 receptor (Herceptin) indicates that targeting a tumor antigen can be useful (Hanna et al. 1999). The approach ideally requires identifying a cell surface protein uniquely expressed on the cells of the tumor but not expressed in the patient's normal tissue exposed to the antibody during therapy. Also promising is a "tumor vaccine" approach where the goal is to direct immune defenses toward the tumor by educating host antigen presenting cells with tumor-derived material (Gilboa et al. 1998). Expression of the marker on the cell surface is not a requirement of this system, but successful systemic administration of a tumor vaccine might require a relative lack of marker expression in all normal tissue cells, especially within vital organs. Either of these therapeutic approaches could benefit from the discovery of new tumor specific markers.

Tumor markers and antigens have promising clinical utility, but previous techniques for locating these proteins have not yielded robust markers for most cancers (Wu 1999). Finding a candidate marker is frequently the by-product of other studies but not the initial intent of the research. Furthermore, generating the expression profile for each suspect gene has often relied on time-consuming techniques, such as northern blotting, in situ hybridization, or immunohisto-

[5]These authors contributed equally to this work.
[6]Corresponding author.
E-MAIL greg.riggins@duke.edu; FAX (919) 681-2796.
Article and publication are at www.genome.org/cgi/doi/10.1101/gr.138000.

chemistry. Fortunately, new genome-scale technology should accelerate tumor marker discovery. In particular, the ability to assay comprehensive gene expression has made significant advances (Gress et al. 1992; Schena et al. 1995; Velculescu et al. 1995; Lockhart et al. 1996; Kononen et al. 1998).

Large-scale gene expression assays, such as cDNA microarrays (Schena et al. 1995), oligonucleotide chips (Lockhart et al. 1996), cDNA library sequencing (Adams et al. 1993), and serial analysis of gene expression (SAGE; Velculescu et al. 1995) can decipher complex expression patterns. Much of the resulting data is being deposited on publicly accessible web sites (Table 1) or is commercially available. Potentially, this information is a valuable resource, but mining the best data and adapting the results for a particular application is challenging. Follow-up and confirmatory studies are time consuming, and this problem will grow with the growth of large-scale expression technologies. A rapid confirmation of differential expression is useful before studies of gene function or before investigating an

overexpressed gene as a candidate tumor marker or antigen.

In this study, we mined a public database for candidate genes (see our previous report on this database; Lal et al. 1999) and used fluorescent-PCR expression comparison (F-PEC) to assess their expression on a panel of tumor and normal samples. The F-PEC method is based on continuous fluorescent monitoring of PCR products (Wittwer et al. 1997; Morrison et al. 1998) from a cDNA template. F-PEC allows for a quick and low-cost assessment of the expression pattern of a gene, uses commercially available instrumentation, and can be automated. From the data obtained, we identified several candidate tumor markers for glioblastoma multiforme (GBM; WHO Astrocytoma Grade IV), which is the most common primary brain malignancy in adults but which can occur at virtually any age (Kleihues et al. 2000). The purpose of this work was to develop the means to find genetic targets specific for GBM that might eventually be useful for developing immune-based therapies. Though we tested our ap-

Table 1. Human Gene Expression Databases

Web site	URL	Tissues	Description
ArrayExpress Database	www.ebi.ac.uk/arrayexpress	Pending	Planned public repository of microarray data
Body Map	bodymap.ims.u-tokyo.ac.jp	Various normal and tumor	Based on EST sequencing of ~3000 reads each from 60 tissues
Brown Lab	cmgm.stanford.edu/pbrown	Cultured fibroblasts and B-cell lymphoma	Custom cDNA array data and supplemental information
CGAP cDNA xProfiler	www.ncbi.nlm.nih.gov/ncicgap/cgapxpsetup.cgi	Most normal and tumor tissues	Can search for cDNA library specific transcripts based on EST sequencing; most extensive profile
CGAP SAGEmap	www.ncbi.nlm.nih.gov/SAGE	Brain, colon ovary, prostate and breast; tumor and normal	Ongoing SAGE expression analysis of cancer; ~50,000 tags from each of 68 tissues and cell lines
Developmental Therapeutics	www.dtp.nci.nih.gov	NCI panel of 60 cancer cell lines	Microarray and drug response data for NCI60 cell lines
Gene Expression Omnibus (GEO)	www.ncbi.nlm.nih.gov/geo	Pending	Future public repository and comparison interface for all types of expression data
Human Gene Expression Index (HuGE)	www.hugeindex.org	Normal blood, kidney, cervix and vulva	Affymetrix chip based database, expanding to cover many normal tissues
Prostate Expression Database	chroma.mbt.washington.edu/PEDB	Prostate normal, premalignant, and cancer	Expression based on 60,000 ESTs in an easily queried interface
SAGEnet (Kinzler-Vogelstein Lab)	www.sagenet.org	Normal colon, colon, and pancreas cancer	Over 400,000 SAGE transcript tags in downloadable files for local analysis
Genexpress-CNRS	idefix.upr420.vjf.cnrs.fr/EXPR/	Brain, muscle, and other normals	5058 brain- or muscle-derived genes probed by filter array
TIGR Human Gene Index	www.tigr.org/tdb/hgi/hgi.html	Most normal tissues	Has query interface for tissue-specific ESTs
Whitehead/MIT Genome Center's Molecular Pattern Recognition Web site	waldo.wi.mit.edu/MPR	Leukemias and hematopoeitic cell lines	Affymetrix based array data available for download

proach on GBM, now that expression information is readily available for many cancerous tissues, aspects of the approach can be employed to help find markers in other major tumors.

RESULTS AND DISCUSSION

Database Mining

Figure 1 outlines the overall experimental procedure we used, starting with database mining for candidate tumor markers. Currently, the Cancer Genome Anatomy Project (CGAP; Strausberg et al. 1997) is the primary public source for gene expression and, in particular, for brain tumors and normal neural tissues (see Table 1). Serial analysis of gene expression (SAGE) data (Velculescu et al. 1995) from CGAP (www.ncbi.nlm. nih.gov/SAGE) was initially chosen for mining EST-based libraries because SAGE libraries are not normal-

Figure 1 Process for finding candidate tumor markers, using fluorescent-PCR expression comparison (F-PEC). Genes overexpressed in tumors are mined from gene-expression databases. A normalized cDNA panel is used to rapidly compare expression levels in malignant and normal tissues. The process requires an initial PCR to determine the specificity of the primers, the product melting temperature, and the expression range of the tested samples. The highest-expressing sample from the first PCR is serially diluted to create a standard curve for a second PCR, yielding information on the relative expression over several orders of magnitude.

ized and because there are significantly more transcript tags available for analysis in SAGE libraries. This predicts a greater sensitivity for detecting low-abundance transcripts in normal tissues. SAGE tags—each representing one transcript—from surgically resected GBM and normal human brain white matter were downloaded, and their numbers were compared. Electronic profiling of transcript numbers revealed 47,500 uniquely expressed neural genes of which 76 genes (0.16%) were overexpressed in the tumors to the order of 10-fold or more and with P values <0.001. From the 76 candidates, 13 genes were chosen for further analysis. Our criteria were that the genes have little or no expression as detected by SAGE in normal brain and that, preferably, they code for cell surface or excreted proteins.

There are other approaches for mining overexpressed genes. For example, a recently developed prediction algorithm for tumor marker discovery from EST data could also improve and supplement the initial candidate selection process (Walker et al. 1999). Many of the Web sites listed in Table 1 have included tools to mine data, and data from these sites can be combined to enhance selection. Regardless of the type or combination of data queried for genes overexpressed in tumors, there is still a need to confirm and expand the expression information. With a rapid confirmation process, candidates from multiple databases can be tested until the genes with the desired pattern of expression are elucidated.

Tissue cDNA Panel

We next sought to verify the SAGE predictions of expression and test expression in a wider range of tissue. In particular, we wanted to know the expression pattern in an independent set of tumors and normal tissue. RNA was extracted from 27 tissues including high-grade astrocytic tumors, normal neural tissues, and normal vital organ tissues. In an attempt to control for varying amounts of cDNA, we first normalized products from cDNA synthesis reactions to β-actin levels. In addition, we checked the total DNA content in these samples using a fluorescent probe with preferential binding to double-stranded nucleic acid. Controls for genomic DNA contamination were all negative. For the 27 samples there was a threefold range of fluorescence indicating that total cDNA amounts varied despite β-actin normalization. Further inspection of β-actin transcript levels in 15 tumor and normal bulk tissues from the SAGEmap database (Lal et al. 1999) showed a 10-fold range of expression, with 377 ± 276 (mean ± SD) transcripts per cell. These results suggested a problem with normalizing to β-actin. SAGE results predicted a tighter control for the s28 ribosomal transcript levels compared with β-actin (Velculescu et al. 1999), but variation in our panel, measured by fluo-
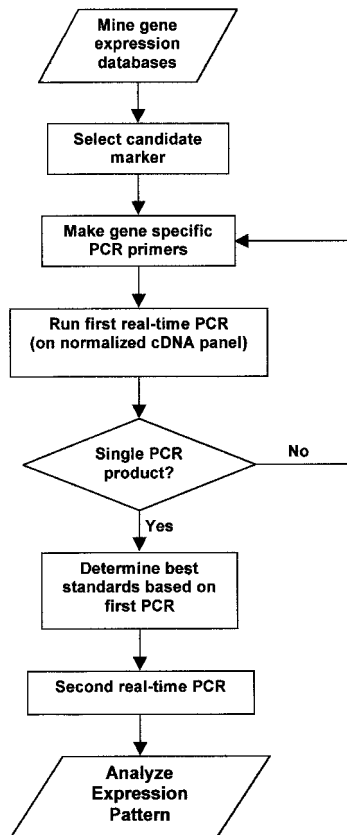
rescent PCR, was not significantly improved over β-actin variation. Such results indicate that normalization to a single housekeeping gene is likely to produce a wide variation in the fractional representation of the target gene from tissues of dissimilar origin. When genes from these tissues are being compared, a better approach should be to normalize by total cDNA levels, with a separate confirmation of the cDNA integrity—by fluorescent PCR or other means—from a housekeeping gene.

To detect candidate tumor antigens, we sought significant expression of the gene in tumor combined with nondetectable expression in normal tissues. Because of the potential problems with normalization and other possible errors, we thought it best to base the decision to proceed with investigating a candidate tumor marker only on absolute differences in expression between tumor and normal tissues and not on small ratios of change.

Fluorescent PCR Verification

Real-time fluorescent PCR has the potential to measure gene expression rapidly in multiple samples and to do so with very sensitive levels of detection (Freeman et al. 1999), capabilities that made evaluating the expression pattern of the mined genes more efficient. PCR primer sets were selected for each candidate gene (Table 2) and optimized for use with the LightCycler (Roche Diagnostics), one of several thermal-cycling machines available that is capable of continuous fluorescence monitoring. One objective was to make the F-PEC procedure rapid and inexpensive, so we avoided the use of fluorescent-labeled hybridization primers. We first tested SYBR green, a fluorescent DNA binding dye with specificity for double-stranded DNA used previously

for this purpose (Morrison et al. 1998). After trying several different PCR-reaction mixtures, the combination of SYBR green, a 'hot-start' type *taq* polymerase, and a modified PCR buffer worked robustly and was relatively inexpensive. To eliminate the potentially confounding effects of primer-dimer amplification, we measured the fluorescence at a temperature below the melting point of the products and above the melting point of the primer-dimers that formed in some reactions. The assay proved to be proportional to the starting cDNA concentrations as determined by serial dilution experiments. Assays using additional fluorescent primers that hybridized within the PCR product (e.g., "taq-man" or fluorescent resonant energy transfer primers) may provide additional assay specificity and sensitivity but might prove difficult to optimize for a rapid screening procedure.

Of 13 genes tested for gene expression levels in our cDNA panel, we were able to quickly find primer pairs for 11 that produced satisfactory PCR amplification for the fluorescent-PCR assay. Conditions were optimized to produce a single PCR product band at the predicted fragment length. Next, the entire normal and tumor tissue panel was assayed to determine the tumor with the highest level of expression. If the initial profile showed increased expression in tumor samples, then the highest expressing tumor was used as a serially diluted standard for a second PCR-based comparison of the sample panel. This second round served as a reproducibility check and ensured that the gene expression levels of all the tissues could be compared simultaneously without extrapolation beyond the standard curve. An outline of the overall approach is shown in Figure 1, and examples of the results are shown in Figure 2. We found the optimization of a gene-specific

Table 2. Candidate Glioblastoma Tumor Markers

Gene name (synonym)	Symbol (GenBank no.)	Cellular location	Primer pairs	$T_a°C^a$
Sec61 gamma	SEC61G (AF054184)	ER-Golgi (Greenfield and High 1999)	TTA CTT TAA TTT AGA AAT AG/ ATC AGG TAA TGC AGT TTG TT	50
Nicotinamide N-methyltransferase	NNMT (U08021)	Cytoplasm (Aksoy et al. 1994)	CTG CCT AGA CGG TGT GAA G/ AGT GGC TGG CTC TGA GTC AC	55
ATP-binding cassette, subfamily C Member 3 (MLP2, MRP3, CMOAT2)	ABCC3 (AB010887)	Membrane (Kiuchi et al. 1998)	CAT CGA CCT GGA GAC TGA CAA C/ CCA TTC TGC GGA CAT ATT TG	58
Neuromedin B (NMB)	NMB (M21551)	Secreted (Krane et al. 1988)	AGC CAG CAA GAT CCG AGT G/ GCA CAA TCT AAG CCA CGC TG	50
Annexin A1 (lipocortin 1, LPC1)	ANXA1 (X05908)	Membrane (Wallner et al. 1986)	GCA GGC CTG GTT TAT TGA AA/ GGT TGC TTC ATC CAC ACC TT	53
SPARC: secreted protein, acidic, cysteine-rich (osteonectin)	SPARC (J03040)	Secreted (Lane and Sage 1994)	AGG TCA CAG GTC TCG AAA A/ AGA GGT GGT GGA AGA AAC TG	53
Glycoprotein (transmembrane) nmb	GPNMB (X76534)	Membrane (Weterman et al. 1995)	AAC TCT ACC CAG TGT GGA AG/ TTG AGG AAG TGG CTA GGA TC	55

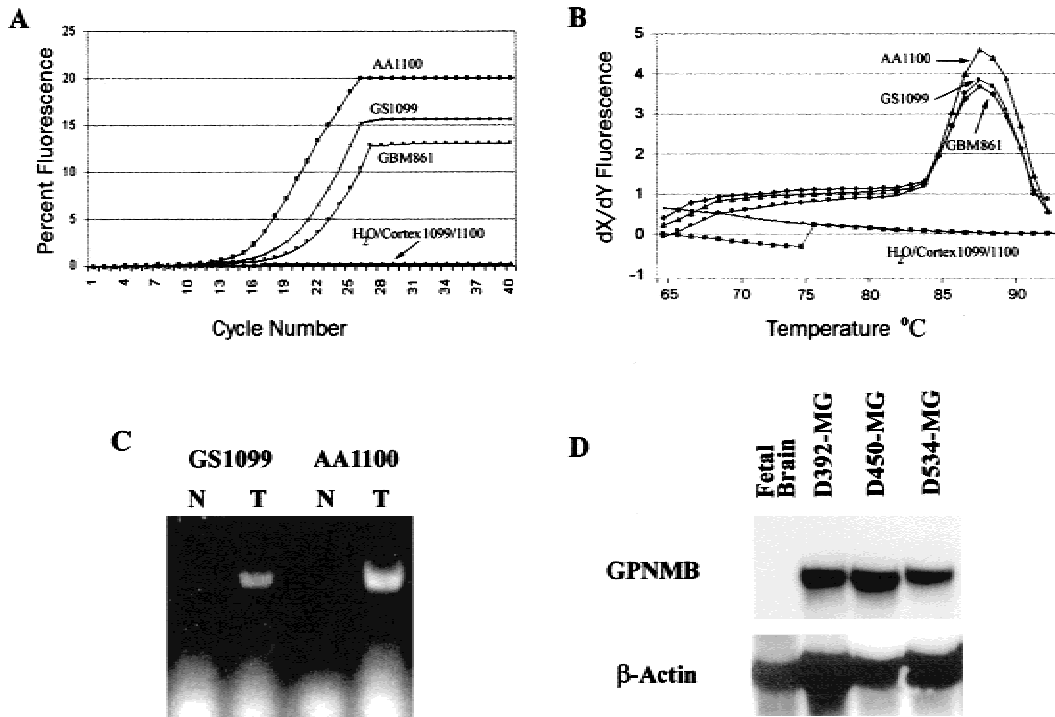$^a T_a$ refers to the annealing temperature used for fluorescent-PCR amplification.

Figure 2 Fluorescent-PCR verification of a candidate glioblastoma marker, *GPNMB*. (*A*) Template cDNA from a bulk glioblastoma (GBM 861) and matched glioma/normal tissue pairs (GS1099/Cortex1099 and AA1100/Cortex1100) were amplified with primers specific for *GPNMB*. (*B*) Melting curve analysis is performed simultaneously to optimize detection temperature, revealing a single peak consistent with a single amplification product. (*C*) After fluorescent-PCR, all reaction products were visualized on an agarose gel to verify a single product of the correct size. (*D*) Northern blot of normal fetal brain and three established GBM cell lines also show a difference in expression for *GPNMB*.

assay to be rapid, requiring the purchase of only one or two unmodified PCR primer pairs per gene. The scheme presented here provides a system that is straightforward to apply with the possibility for higher throughput automation.

A significant problem with any gene expression assay is assessing the purity of the samples tested. Primary tumor tissue has varying degrees of normal cells, and "normal" tissue obtained during tumor resection may have occult malignant cells. For example, infiltration by macrophages into the tumor samples might produce a marker against a nonmalignant cell population. To guard against this latter possibility, we relied on additional expression in well-established glioblastoma cell lines, presumed to be a purely malignant cell population, for both the candidate selection and F-PEC analysis. When high expression of these genes was observed in a GBM cell line, it suggested that expression in the bulk tumor was from transformed cells and not from normal cells.

Another potential problem with this approach is unrecognized gene expression in a small but vital normal cell population. Assaying a greater range of normal tissues or defined cell populations can perhaps minimize this. Certainly, the existence of a desirable expression pattern in a potential tumor marker is only sug-

gestive of its potential as a truly useful marker. Further immunohistochemical or in situ hybridization of tissue sections will be required on a culled set of the most promising tumor marker candidates. Since development of novel antibodies is time consuming and expensive, the F-PEC approach may be useful in triaging candidates before antibody design and synthesis. In addition, F-PEC could be readily applied to laser-captured microdissected cells to ensure a greater level of sample purity (Simone et al. 1998).

Of the 11 candidate genes assayed, four were deemed unacceptable due to either high level of expression in one of the normal neural-derived tissues (*CSRP2*, *S100A4*, and *CXCR4*) or was expressed in only one tumor from the panel (*GCS1*). Seven genes showed a distinct difference in transcript levels between normal neural tissues and some GBMs (Fig. 3).

Though genes that have a promising pattern of RNA expression can be found using this procedure, several errors are inherent with this approach. PCR-based assays may suffer from sequence variations at the primer sites, differential splicing, or spurious amplification from related cDNA sequences. In the course of this study, we detected one instance where there was an inconsistency between the Northern blot result and the F-PEC results. There was not consistent PCR ampli-
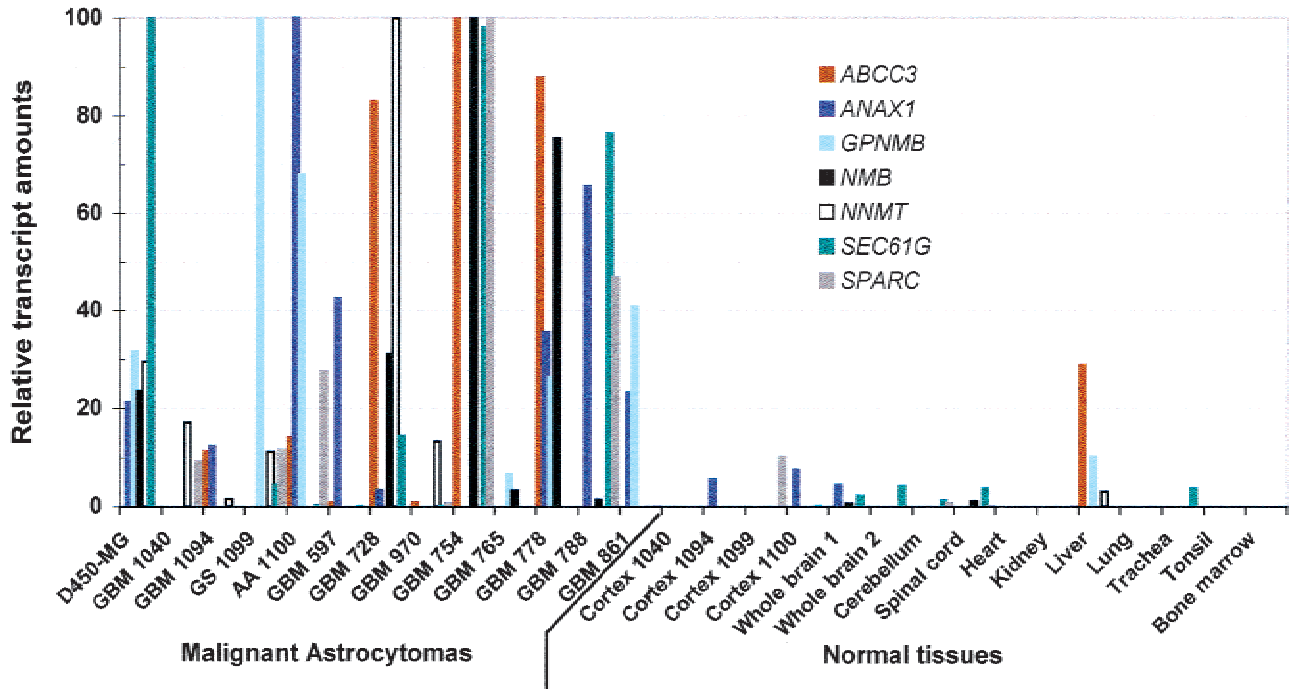
Figure 3 Relative expression of the tumor markers in 12 high-grade astrocytomas, one glioblastoma cell line (D450-MG), and normal tissues. Glioblastoma (GBM), gliosarcoma (GS), anaplastic astrocytoma (AA), and cortex samples with the same patient number indicate matched normal/tumor pairs removed during the same surgery. Gene expression levels determined by fluorescent-PCR were plotted relative to the highest expression tumor in each case. Gene expression is graphically displayed relative to serial dilutions of the highest expressing tumor.

fication in a cell line where a band was observed on a Northern blot but not amplified by PCR. We hypothesized that our 1.4-kb PCR product length was too long to amplify consistently. Concordant results between the two methods for all samples were achieved after spacing the PCR primers closer together. Based on this observation, our recommendation is that amplification products be designed smaller than ~300 bp for the F-PEC procedure.

Western Blotting

Though high levels of RNA transcripts can be predictive of high protein levels, ultimately protein levels must be confirmed if targeting the tumor antigen is the desired endpoint. Commercial antibodies were available for Annexin A1 and used for Western blotting (Fig. 4). Strong reactivity was observed for GBM cell lines and most of the GBM bulk samples. Compared with a GBM positive control, normal cortex removed from an area adjacent to seizure foci, rapid autopsy cortex, cerebellum, and thalamus samples—all removed from patients without brain tumors—contained little or no detectable protein. One of four tissue samples initially diagnosed as normal cortex adjacent to a GBM was reactive for Annexin A1 (not shown) and may be contaminated with tumor cells as none of the six normal samples from non–cancer patients had detectable protein. The observation of elevated Annexin A1 protein
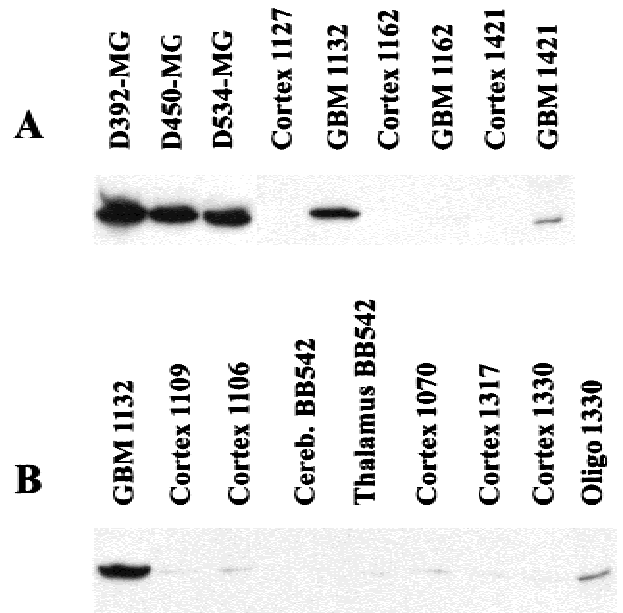

Figure 4 Western blotting of annexin A1. Protein levels from brain tumors, glioblastoma cell lines, and normal neural tissue were compared. (*A*) Bands in both glioblastoma cell lines (D392-MG, D450-MG, and D534-MG) and primary GBM indicate that protein is expressed in transformed, but not normal (Cortex 1127, Cortex 1162, and Cortex 1421) tissues. (*B*) Normal tissues from different normal brain regions did not express high levels of ANXA1 protein compared to a glioblastoma (GBM 1132) or an oligodendroglioma (Oligo 1330).

levels in cancer is consistent with immunohistochemistry revealing reactivity in breast carcinomas but not in normal breast tissue or most benign breast tumors (Ahn et al. 1997).

GBM Tumor Markers

It is possible that no single gene would be up-regulated in all GBM if these tumors arose through different molecular mechanisms. GBM are a heterogeneous group of tumors, with at least two distinct molecular genetic pathways (Kleihues and Ohgaki 1999). The seven selected genes showed a significant increase in expression, on average, in only three of 12 of the tumors assayed. The composite expression showed that eight of 12 glioblastomas had at least one marker that would discriminate between tumor and normal, with a difference in expression of at least 10-fold. The four remaining tumors with no marker expression may have similar histology, but the tumors were molecularly different, at least for these genes, from the original tumors used for the SAGE analysis that showed expression of these genes. Using the set of tumors with no candidate markers for further SAGE analysis, or selecting candidates from different databases, would perhaps yield markers specific to the remaining tumors. These or other tumor markers may eventually provide a means to distinguish between different subclasses of GBM.

The fact that three-fourths of the tumors had at least one gene overexpressed suggests a custom approach to therapy whereby multiple candidate markers in a tumor biopsy are assayed to detect the best combination prior to therapy. Approaches might include injecting toxic antibodies (Panchal 1998) or immunizing a patient's dendritic cells with the RNA from a specific set of tumor markers (Avigan 1999; Bjorck 1999). Therapies applied locally, that is, within the CNS compartment, have an advantage because they may still be useful even if there is gene expression in a distant normal tissue that does not come in contact with the therapy.

On the basis of the tumor-specific expression pattern of several of the genes we tested, if this pattern is maintained at the protein level, applications for these genes may eventually be found. An ATP-binding cassette, subfamily C Member 3 (ABCC3) protein, which has homology to multidrug resistance-associated proteins (Kool et al. 1999) showed the highest induction over normal brain samples. ABCC3 is a transmembrane protein and is therefore a potential target for antibody therapy. However, expression of *ABCC3* was observed in normal liver tissue, which would not make this gene a good target for systemic therapies but perhaps make it useful for localized central nervous system targeting of GBM. For other targets, the possibility of insignificant expression in vital tissue remains, making these genes candidates for systemic therapy pending further testing.

Four of the seven potential glioblastoma tumor markers were previously implicated in cancer and had patterns of expression that would be consistent with overexpression in cancer. Neuromedin B, a bombesin-like growth peptide, is speculated to be an autocrine growth factor for lung cancer (Siegfried et al. 1999) but is likely expressed in normal anterior pituitary (Houben et al. 1993). SPARC, an extracellular matrix protein involved in tissue remodeling, is angiogenic (Jendraschak and Sage 1996) and is implicated in a number of different tumor types, including brain tumors (Ledda et al. 1997; Rempel et al. 1998, 1999). *ABCC3* is overexpressed in various cancer cell lines (Kool et al. 1999) and confers resistance to chemotherapy (Zeng et al. 1999). *Annexin A1* is expressed in gastric cancers and breast carcinoma and is speculated to have immunosuppressive properties important for avoiding a host response to the tumor (Sakata et al. 1993; Ahn et al. 1997; Koseki et al. 1997). Annexin A1 has also been implicated in metastasis of breast adenocarcinomas (Pencil and Toth 1998).

Another approach likely to enhance tumor marker discovery is tissue microarray technology. Tissue microarrays can simultaneously probe expression in hundreds or thousands of tissue cores (Kononen et al. 1998; Moch et al. 1999). F-PEC data could augment tissue microarray analysis, in particular when an antibody or in situ hybridization assay is not readily available for a particular gene. Regardless of the follow-up approach, there is a real need to be able to rapidly assess well-documented samples for expression of genes initially identified by comprehensive gene expression technologies.

The rapid growth of on-line information presents a new challenge to the experimental biologist. How does one efficiently adapt these data for practical applications? Here we have attempted to enhance tumor marker discovery by using public gene expression data followed by rapid expression screening to locate candidate tumor markers for GBM. This study is not exhaustive, searching for all the possible database-mined candidate genes, and it produces only patterns of RNA expression that are suggestive of utility. However, it does indicate that there are some genes that are highly expressed in a portion of GBM but not in surrounding normal neural tissue. These data also suggest that there is no one highly expressed gene common to all tumors classified by histology as GBM. Still, the possibility remains that a combination of genes identified by this approach may eventually be useful for therapy or prognosis. Continued application of F-PEC to the increasing amount of large-scale expression data should yield additional tumor marker candidates. This approach can be easily adapted and applied to various tumor

types, in particular to test candidate genes mined from public databases.

METHODS

Data Mining

Differentially expressed transcript targets were chosen from SAGE data housed at the CGAP Web site (http://www.ncbi.nlm.nih.gov/SAGE/) as previously described (Lal et al. 1999). SAGE tags from four bulk tissue libraries (SAGE_Duke_GBM_H1110, SAGE_pooled_GBM, SAGE_BB542, and SAGE_normal_pool) were downloaded from this site and compared for fold induction and statistical significance using the SAGE 300 program kindly provided by K. Kinzler (see http://www.sagenet.org/). Significance was based on Monte Carlo simulations from this program, with a cut-off at *P*-chance = 0.001. The SAGE libraries were made from two normal brain white matter libraries compared with two GBM bulk tumor libraries, with further details regarding each tissue sample located at the SAGEmap Web site (Table 1). Because of pooling of samples in some of the original SAGE libraries, our comparison reflected transcript levels of RNA derived from three normal samples compared to tumor-derived RNA from six patients. Candidate selection was based on consideration of their relative fold induction in GBM compared with normal brain, lack of predicted expression in normal tissues, expression in the GBM cell line, and for some cases, a known membrane or extracellular localization of the protein.

Sample Descriptions

Normally discarded tumor tissue was snap frozen immediately after surgery and diagnosis and was stored at −135°C. Final pathologic diagnosis of primary bulk tissues used for F-PEC confirmed 11 GBM (Grade IV astrocytomas) and one Grade III anaplastic astrocytoma (AA 1100). One gliosarcoma (GS 1099) is from a class of GBM variants that accounts for ~2% of glioblastomas (Kleihues et al. 2000). The cell lines (D392-MG, D450-MG, and D534-MG) used for F-PEC, Northern blotting, and Western blotting were well established (>70 passages) and originally from an independent set of confirmed GBMs. There was one gliosarcoma variant within this set as well (D392-MG). Tumor tissue for Western blotting was also confirmed through clinical pathology as WHO grade IV glioblastoma multiforme (GBM1132, GBM1162, GBM1421, and GBM1330), plus one WHO grade II well-differentiated oligodendroglioma.

Eight samples were used to represent normal neural tissues for the F-PEC panel: four normal cortex samples that were adjacent to four of the above tumors and RNA purchased from two different adult whole brains, spinal cord, or cerebellum (Clontech). Histology from one margin of ~5-mm pieces of tissue obtained during tumor resection was used to identify the 'normal' brain samples derived from brain tumor patients. Nonneural tissues were also procured in the above fashion either commercially (heart, kidney, liver, and lung) or from autopsy tissue (tonsil, bone marrow, and trachea). For Northern blotting, the normal RNA shown in Figure 2D was fetal brain total RNA (Clontech). For Western blotting, normal tissue was also obtained from a noncancer seizure patient (Cortex 1109, 1106, and 1070) and rapid autopsy samples from normal individuals (cerebellum BB542 and Thalamus BB542).

RNA Isolation and cDNA Synthesis

Total RNA was isolated by separation on a cesium chloride gradient, and messenger RNA was purified from total using oligo-dT cellulose columns (New England Biolabs). Equal amounts of mRNA, as determined by RiboGreen fluorescence (Molecular Probes), were used in identical cDNA synthesis reactions (Superscipt II; Life Technologies, GIBCO). The resulting cDNAs were screened for genomic contamination using genomic specific primers as well as confirming no amplification from control samples lacking reverse-transcriptase. All cDNA samples that lacked any detectable genomic DNA were then normalized to their cDNA concentrations as determined by PicoGreen (Molecular Probes) binding fluorescence.

Northern and Western Blotting

For Northern blot analysis, total RNA was isolated by CsCl ultracentrifugation. Hybridization probes were amplified from target gene sequences and β-actin. Equal amounts of total RNA, as determined by ultraviolet spectrophotometry, were separated on an agarose gel and blotted overnight before hybridizing them with a radioactively labeled PCR product.

For Western analysis, total cell lysates were prepared from corresponding cell pellets and frozen tissue samples. Equal amounts of protein from each sample were separated by electrophoresis and transferred to nitrocellulose membrane. Human Annexin A1-detecting antibody (Transduction Laboratories) was incubated with the membrane for 1 hr, followed by subsequent incubation with horseradish peroxidase-conjugated sheep antimouse immunoglobulin. Protein was visualized by chemiluminescence (Amersham) and exposed to Kodak X-ray film for 5–10 sec. The molecular weights were determined by prestained standards (Life Technologies). Equivalent protein loadings were verified by staining the gel with Comassie blue after transfer.

Fluorescent-PCR Verification

Fluorescent-PCR was performed using a thermocycler (Light-Cycler; Roche Diagnostics) with continuous monitoring of SYBR Green I (Molecular Probes) fluorescence (Morrison et al. 1998) and normalized cDNA templates. The PCR reaction conditions, modified from those previously described (Vogelstein and Kinzler 1999), were 67 mM Tris (pH 8.8), 16.6mM NH_4SO_4, 6.7 mM $MgCl_2$, 10 mM β-mercaptoethanol, 0.5 µg/µl BSA, 1 µl of SYBR green diluted 1 : 1500, 0.25 µM of each PCR primer, 200 µM of each dNTP, and 1 U of platinum taq (Life Technologies) in a final volume of 20 µl.

The integrity of each sample was confirmed using primers specific for β-actin (5′-CGT CTT CCC CTC CAT CG and 5′-CTC GTT AAT GTC ACG CAC). A test of optimal annealing conditions, as well as melting curve analysis, was conducted for each set of gene-specific primers. This allowed us to refine PCR kinetics and conditions for each primer pair and to set the temperature for fluorescence reading between the melting temperature of any primer-dimer formation and the intended amplification product.

A 32-capillary sample rotor for the themocycler was filled for each target assay, permitting an H_2O control, positive control dilutions (to create a standard curve), an independent cell-line positive control, and 27 test samples. The expression levels for each transcript were assayed in 12 primary tumors, six normal brain samples, and nine other normal samples from vital organs. First-round assays were conducted to establish expression levels in normal and tumor tissue. Second-

round reactions were conducted on each cDNA target using dilutions of the highest-expressing tissue (determined from the first run) to compare relative expression of all samples without extrapolation beyond the standard curve. This additional run also served as a check of reproducibility. Fluorescence curves obtained from the LightCycler system were analyzed by a second derivative fit for quantification analysis of transcript targets. The second derivative method used the point for which the rate of change of fluorescence is maximized, created a fit to the log-linear portion of the amplification curve, and extrapolated the starting concentration. Relative expression was determined by comparison to three control samples serially diluted 10-fold. After each assay, the reaction mixture was run on an agarose gel to visualize results and to verify a single band of the correct size.

ACKNOWLEDGMENTS

REFERENCES

Adams, M.D., Soares, M.B., Kerlavage, A.R., Fields, C., and Venter, J.C. 1993. Rapid cDNA sequencing (expressed sequence tags) from a directionally cloned human infant brain cDNA library. *Nat. Genet.* **4:** 373–380.

Ahn, S.H., Sawada, H., Ro, J.Y., and Nicolson, G.L. 1997. Differential expression of annexin I in human mammary ductal epithelial cells in normal and benign and malignant breast tissues. *Clin. Exp. Metastasis* **15:** 151–156.

Aksoy, S., C.L. Szumlanski, and R.M. Weinshilboum. 1994. Human liver nicotinamide N-methyltransferase: cDNA cloning, expression, and biochemical characterization. *J. Biol. Chem.* **269:** 14835–14840.

Avigan, D. 1999. Dendritic cells: Development, function and potential use for cancer immunotherapy. *Blood Rev.* **13:** 51–64.

Bjorck, P. 1999. Development of dendritic cells and their use in tumor therapy. *Clin. Immunol.* **92:** 119–127.

Freeman, W.M., Walker, S.J., Vrana, K.E., Orlando, C., Pinzani, P., and Pazzagli, M. 1999. Quantitative RT-PCR: Pitfalls and potential developments in quantitative PCR. *Biotechniques* **26:** 112–122, 124–125.

Gilboa, E., Nair, S.K., and Lyerly, H.K. 1998. Immunotherapy of cancer with dendritic-cell-based vaccines. *Cancer Immunol. Immunother.* **46:** 82–87.

Greenfield, J.J. and High, S. 1999. The Sec61 complex is located in both the ER and the ER-Golgi intermediate compartment. *J. Cell Sci.* **112:** 1477–1486.

Gress, T.M., Hoheisel, J.D., Lennon, G.G., Zehetner, G., and Lehrach, H. 1992. Hybridization fingerprinting of high-density cDNA-library arrays with cDNA pools derived from whole tissues. *Mamm. Genome* **3:** 609–619.

Hanna, W., Kahn, H.J., and Trudeau, M. 1999. Evaluation of HER-2/neu (erbB-2) status in breast cancer: From bench to bedside. *Mod. Pathol.* **12:** 827–834.

Houben, H., Vandenbroucke, A.T., Verheyden, A.M., and Denef, C. 1993. Expression of the genes encoding bombesin-related peptides and their receptors in anterior pituitary tissue. *Mol. Cell. Endocrinol.* **97:** 159–164.

Jendraschak, E. and Sage, E.H. 1996. Regulation of angiogenesis by SPARC and angiostatin: Implications for tumor cell biology. *Semin. Cancer Biol.* **7:** 139–146.

Kardamakis, D. 1996. Tumour serum markers: Clinical and economical aspects. *Anticancer Res.* **16:** 2285–2288.

Kiuchi, Y., Suzuki, H., Hirohashi, T., Tyson, C.A., and Sugiyama, Y. 1998. cDNA cloning and inducible expression of human multidrug resistance associated protein 3 (MRP3). *FEBS Lett.* **433:** 149–152.

Kleihues, P., Burger, P.C., Plate, K.H., Ohgaki, H., and Cavenee, W.K. 2000. *Pathology & genetics: Tumours of the nervous system* (ed. P. Kleihues and W.K. Cavenee). International Agency for Research on Cancer, Lyon.

Kleihues, P. and Ohgaki, H. 1999. Primary and secondary glioblastomas: From concept to clinical diagnosis. *Neuro-Oncology* **1:** 44–51.

Kononen, J., Bubendorf, L., Kallioniemi, A., Barlund, M., Schraml, P., Leighton, S., Torhorst, J., Mihatsch, M.J., Sauter, G., and Kallioniemi, O.P. 1998. Tissue microarrays for high-throughput molecular profiling of tumor specimens. *Nat. Med.* **4:** 844–847.

Kool, M., van der Linden, M., de Haas, M., Scheffer, G.L., de Vree, J.M., Smith, A.J., Jansen, G., Peters, G.J., Ponne, N., Scheper, R.J., et al. 1999. MRP3, an organic anion transporter able to transport anti-cancer drugs. *Proc. Natl. Acad. Sci.* **96:** 6914–6919.

Koseki, H., Shiiba, K., Suzuki, Y., Asanuma, T., and Matsuno, S. 1997. Enhanced expression of lipocortin-1 as a new immunosuppressive protein in cancer patients and its influence on reduced in vitro peripheral blood lymphocyte response to mitogens. *Surg. Today* **27:** 30–39.

Krane, I.M., Naylor, S.L., Helin-Davis, D., Chin, W.W., and Spindel, E.R. 1988. Molecular cloning of cDNAs encoding the human bombesin-like peptide neuromedin B. Chromosomal localization and comparison to cDNAs encoding its amphibian homolog ranatensin. *J. Biol. Chem.* **263:** 13317–133123.

Lal, A., Lash, A.E., Altschul, S.F., Velculescu, V., Zhang, L., McLendon, R.E., Marra, M.A., Prange, C., Morin, P.J., Polyak, K., et al. 1999. A public database for gene expression in human cancers. *Cancer Res.* **59:** 5403–5407.

Lane, T.F. and Sage, E.H. 1994. The biology of SPARC, a protein that modulates cell-matrix interactions. *Faseb J.* **8:** 163– 173.

Ledda, M.F., Adris, S., Bravo, A.I., Kairiyama, C., Bover, L., Chernajovsky, Y., Mordoh, J., and Podhajcer, O.L. 1997. Suppression of SPARC expression by antisense RNA abrogates the tumorigenicity of human melanoma cells. *Nat. Med.* **3:** 171–176.

Lockhart, D.J., Dong, H., Byrne, M.C., Follettie, M.T., Gallo, M.V., Chee, M.S., Mittmann, M., Wang, C., Kobayashi, M., Horton, H., et al. 1996. Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat. Biotechnol.* **14:** 1675–1680.

Lokeshwar, V.B., Obek, C., Soloway, M.S., and Block, N.L. 1997. Tumor-associated hyaluronic acid: A new sensitive and specific urine marker for bladder cancer. *Cancer Res.* **57:** 773–777.

Mao, L., Lee, D.J., Tockman, M.S., Erozan, Y.S., Askin, F., and Sidransky, D. 1994. Microsatellite alterations as clonal markers for the detection of human cancer. *Proc. Natl. Acad. Sci.* **91:** 9871–9875.

Moch, H., Schraml, P., Bubendorf, L., Mirlacher, M., Kononen, J., Gasser, T., Mihatsch, M.J., Kallioniemi, O.P., and Sauter, G. 1999. High-throughput tissue microarray analysis to evaluate genes uncovered by cDNA microarray screening in renal cell carcinoma. *Am. J. Pathol.* **154:** 981–986.

Morrison, T.B., Weis, J.J., and Wittwer, C.T. 1998. Quantification of low-copy transcripts by continuous SYBR Green I monitoring during amplification. *Biotechniques* **24:** 954–958, 960, 962.

Panchal, R.G. 1998. Novel therapeutic strategies to selectively kill cancer cells. *Biochem. Pharmacol.* **55:** 247–252.

Pencil, S.D. and Toth, M. 1998. Elevated levels of annexin I protein in vitro and in vivo in rat and human mammary adenocarcinoma. *Clin. Exp. Metastasis* **16:** 113–121.

Rempel, S., Ge, S., and JA, G. 1999. SPARC: A potential diagnostic marker of invasive meningiomas. *Clin. Cancer Res.* **5:** 237–241.

Rempel, S.A., Golembieski, W.A., Ge, S., Lemke, N., Elisevich, K.,

Mikkelsen, T., and Gutierrez, J.A. 1998. SPARC: A signal of astrocytic neoplastic transformation and reactive response in human primary and xenograft gliomas. *J. Neuropathol. Exp. Neurol.* **57:** 1112–1121.

Sakata, T., Iwagami, S., Tsuruta, Y., Suzuki, S., and Suzuki, R. 1993. Study of natural lipocortin. I. A potent mediator for macrophage-mediated immunosuppression in tumor-bearing mice. *J. Immunol.* **151:** 4964–4972.

Schena, M., Shalon, D., Davis, R.W., and Brown, P.O. 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270:** 467–470.

Sidransky, D., Tokino, T., Hamilton, S.R., Kinzler, K.W., Levin, B., Frost, P., and Vogelstein, B. 1992. Identification of ras oncogene mutations in the stool of patients with curable colorectal tumors. *Science* **256:** 102–105.

Siegfried, J.M., Krishnamachary, N., Gaither Davis, A., Gubish, C., Hunt, J.D., and Shriver, S.P. 1999. Evidence for autocrine actions of neuromedin B and gastrin-releasing peptide in non-small cell lung cancer. *Pulm. Pharmacol. Ther.* **12:** 291–302.

Simone, N.L., Bonner, R.F., Gillespie, J.W., Emmert-Buck, M.R., and Liotta, L.A. 1998. Laser-capture microdissection: Opening the microscopic frontier to molecular analysis. *Trends Genet.* **14:** 272–276.

Strausberg, R.L., Dahl, C.A., and Klausner, R.D. 1997. New opportunities for uncovering the molecular basis of cancer. *Nat. Genet.* **15:** 415–416.

Velculescu, V.E., Madden, S.L., Zhang, L., Lash, A.E., Yu, J., Rago, C., Lal, A., Wang, C.J., Beaudry, G.A., Ciriello, K.M., et al. 1999. Analysis of human transcriptomes. *Nat. Genet.* **23:** 387–388.

Velculescu, V.E., Zhang, L., Vogelstein, B., and Kinzler, K.W. 1995. Serial analysis of gene expression. *Science* **270:** 484–487.

Vogelstein, B. and Kinzler, K.W. 1999. Digital PCR. *Proc. Natl. Acad. Sci.* **96:** 9236–9241.

Walker, M.G., Volkmuth, W., Sprinzak, E., Hodgson, D., and Klingler, T. 1999. Prediction of gene function by genome-scale expression analysis: Prostate cancer-associated genes. *Genome Res.* **9:** 1198–1203.

Wallner, B.P., Mattaliano, R.J., Hession, C., Cate, R.L., Tizard, R., Sinclair, L.K., Foeller, C., Chow, E.P., Browing, J.L., Ramachandran, K.L., et al. 1986. Cloning and expression of human lipocortin, a phospholipase A2 inhibitor with potential anti-inflammatory activity. *Nature* **320:** 77–81.

Weterman, M.A., Ajubi, N., van Dinter, I.M., Degen, W.G., van Muijen, G.N., Ruitter, D.J., and Bloemers, H.P. 1995. nmb, a novel gene, is expressed in low-metastatic human melanoma cell lines and xenografts. *Int. J. Cancer* **60:** 73–81.

Wittwer, C.T., Herrmann, M.G., Moss, A.A., and Rasmussen, R.P. 1997. Continuous fluorescence monitoring of rapid cycle DNA amplification. *Biotechniques* **22:** 130–131, 134–138.

Wu, J.T. 1999. Review of circulating tumor markers: From enzyme, carcinoembryonic protein to oncogene and suppressor gene. *Ann. Clin. Lab. Sci.* **29:** 106–111.

Zeng, H., Bain, L.J., Belinsky, M.G., and Kruh, G.D. 1999. Expression of multidrug resistance protein-3 (multispecific organic anion transporter-D) in human embryonic kidney 293 cells confers resistance to anticancer agents. *Cancer Res.* **59:** 5964–5967.