Gene Expression Profiling from Formalin-Fixed Paraffin-Embedded Tumors of Pediatric Glioblastoma

Takrima Haque,[1] Damien Faury,[1] Steffen Albrecht,[2] Enrique Lopez-Aguilar,[6] Péter Hauser,[7] Miklós Garami,[7] Zoltán Hanzély,[9] László Bognár,[8] Rolando F. Del Maestro,[4] Jeffrey Atkinson,[3] Andre Nantel,[5] and Nada Jabado[1]

Abstract

Purpose: Gene expression profiling has proved crucial for understanding the biology of cancer. In rare diseases, including pediatric glioblastoma (pGBM), the lack of readily available fresh frozen (FF) material limits the feasibility of this analysis, as well as its validation, on independent data sets, a step needed to ensure relevance, mandating the use of alternate RNA sources. To overcome the limitation of material number and to validate results we obtained on FF pGBM, we did microarray analysis on RNA extracted from formalin-fixed, paraffin-embedded archival samples from pGBM and control brains, wherein we had no control on the fixation process.

Experimental Design: RNA from 16 pGBM and 3 control brains was extracted and linearly amplified. Reverse transcription – PCR on housekeeping and formerly identified tumor-associated genes and microarray analysis were done on this RNA source. Results were validated by immunohistochemistry.

Results: Despite extensive RNA degradation, microarray analysis was possible on 16 of 19 samples and reproduced the pattern of results obtained on FF pGBM. Gene lists and ontology subgrouping were highly concordant in both sample types. Similar to the findings on FF samples, we were able to identify two subsets of pGBM based on their association/lack of association with evidence consistent with an active Ras pathway.

Conclusions: Archival formalin-fixed, paraffin-embedded tissues are an invaluable resource as they are the most widely available materials often accessible in conjunction with clinical and follow-up data. Gene expression profiling on this material is feasible and may represent a significant advance for understanding the biology of rare human diseases.

In the last decade, gene expression profiling of human cancer has proved valuable in cancer research, providing precious insight into mechanisms and targets involved in oncogenesis in several neoplasms (1, 2). The applications of microarray and transcript profiling analysis have, however, been limited by the need for fresh frozen (FF) materials, which allow the extraction of high-quality nucleic acids, and by the limited clinical and outcome data associated with available FF tissues. On the other side, the collection and storage of archival, formalin-fixed and paraffin-embedded (FFPE) tissue specimens needed to establish diagnosis is a routine practice in pathology laboratories. Moreover, this sample type is often available in conjunction with precious clinical and follow-up data (3). However, extraction of quality RNA for gene expression analysis from FFPE samples has proved difficult because of the extensive degradation and fragmentation that occurs during the fixation process. There is cross-linkage between the nucleic acids and the proteins that covalently modifies the RNA, making subsequent expression analysis a technical challenge (4). Several groups showed reliable and reproducible results for reverse transcription – PCR (RT-PCR) and quantitative RT-PCR (qRT-PCR; refs. 5 – 12). However, multiple gene expression analysis has proved more challenging, with few reports available that mainly address qRT-PCR on multiple gene sets (13 – 18). For most of these studies, investigators had control on the type of fixative and the process, whereas for others, the validation of transcript profiles obtained from FF versus this material source has been problematic based on the difficulty in obtaining matched frozen and fixed samples (13).

Brain tumors are the leading cause of cancer-related mortality in children. Pediatric glioblastoma (pGBM; pediatric grade IV

Authors' Affiliations: [1]Division of Hematooncology, Department of Pediatrics, [2]Department of Pathology, [3]Division of Neurosurgery, Montreal Children's Hospital and [4]Division of Neurosurgery, Montreal Neurological Institute, Brain Tumour Research Centre, McGill University Health Center; [5]Biotechnology Research Institute, National Research Council of Canada, Montreal, Canada; [6]Oncology Department, Pediatrics Hospital, Centro Medico Nacional Siglo XXI, Mexico City, Mexico; [7]2nd Department of Pediatrics, Faculty of Medicine, Semmelweis University; [8]Division of Neurosurgery, Division of Pathology, National Institute of Neurosurgery, Budapest, Hungary; and [9]Department of Neurosurgery, Medical and Health Science Center, University of Debrecen, Debrecen, Hungary

Received 3/5/07; revised 7/9/07; accepted 7/25/07.

Grant support: Canadian Institute of Health Research and the Penny Cole Foundation (NJ), the NRC Genome Health Initiative (AN), the Hungarian Scientific Research Fund (OTKA) Contract No. T-04639, and the National Research and Development Fund (NKFP) Contract No. 1A/002/2004 (PH, MG, LB, ZH). N. Jabado is the recipient of a Chercheur Boursier Award from Fonds de la Recherche en Sante du Quebec. This is NRC publication number 49512.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Requests for reprints: Nada Jabado, Montreal Children's Hospital Research Institute, 4060 Saint Catherine West, PT-239, Montreal, Quebec, Canada, H3Z 2Z3. Phone: 514-412-4400, ext. 23270; Fax: 514-412 4331; E-mail: nada.jabado@mcgill.ca.

doi:10.1158/1078-0432.CCR-07-0525

astrocytoma) is a rare and deadly brain tumor (19–22). It accounts for 15% of all pediatric brain tumors, has a 3-year survival of <20%, and is associated with high morbidity (23). Considerable information is available on adult GBM (19–21, 24); however, fewer molecular data exist on the mechanisms underlying its development in children, mainly because of the relative lack of frozen samples (25, 26). We previously generated gene expression profiles on a set of 14 pGBM FF samples using the University Health Network human 19K cDNA microarrays (27). In the absence of readily available FF pGBM samples to validate our microarray results on an independent data set from the same tumor, we considered the use of other more readily available RNA sources, such as RNA from FFPE samples. Our aim was to assess if archival materials can be used to investigate expression profiles of this tumor and other rare tumors that lack readily available FF material and clinical data.

Materials and Methods

Sample characteristics and pathologic review. A senior neuropathologist, Dr. S. Albrecht, reviewed all samples included in this study to ensure consistent classification based on contemporary guidelines from WHO. Only GBMs were examined. All samples were obtained with informed consent after approval of the institutional review board of the respective hospitals they were treated in. Seventeen FF samples, including 14 pGBM samples and 3 pediatric control brains (CB), and 19 FFPE samples, including 16 pGBM and 3 pediatric CB, from children of ages 3, 5, and 14 years, respectively, were processed after central review (Table 1). All of the FF and FFPE samples came from distinct individuals (clinical findings summarized in Table 1). All of the six CBs were obtained from surgical procedures on pediatric patients with epilepsy or congenital malformations and were reviewed by the neuropathologist to ascertain for astrocytic content. Tissues were obtained from Pediatric Cooperative Human Tissue Network, London/Ontario Tumor Bank, and collaborators in Montreal, Hungary, and Mexico. All FFPE blocks were collected >3 years before RNA extraction and analyses.

RNA extraction after scrape or laser capture microdissection. Seven-micrometer sections from FFPE blocks were collected onto precleaned Superfrost Plus slides (VWR Scientific), stored at -70°C, and used within 15 days. A representative slide for each sample was stained with H&E, and tumor areas were identified by the neuropathologist. Sections were processed following manufacturer's instructions (Paradise kit, Arcturus) and as described previously (27, 28). FF blocks were

Table 1. Characteristics of the patients included in the study

FFPE pGBM samples	Gender	Age (y)	Ras (IHC)	YB1 (IHC)	Survival from surgery
P1	M	5	pos	pos	D, 7 mo
P2	F	7	neg	neg	A, 4 y
P3	F	16	neg	neg	D, 12 mo
P4	M	4	pos	pos	D, 3 y
P5	M	14	neg	neg	D, 15 mo
P6	F	12	pos	pos	D, 11 mo
P7	M	9	pos	pos	D, 14 mo
P8	M	14	pos	pos	D, 2.5 y
P9	M	8	pos	pos	D, 9 mo
P10	F	8	pos	pos	D, 14 mo
P11	F	12	pos	pos	D, 3 y
P12	F	7	pos	pos	D, 2 mo
P13	F	9	pos	pos	D, 36 mo
P14	M	4	pos	pos	D, 2 y
P15	M	12	neg	neg	A, 6 y
P16	F	15	neg	neg	D, 2 y
CB1	M	3	neg	neg	A
CB2	M	5	neg	neg	A
CB3	F	14	neg	neg	A
FF pGBM samples	**Gender**	**Age (y)**	**Ras (WB)**	**YB1 (qRT-PCR)**	**Survival from surgery**
FF1	F	7	neg	1.8	A, 4 y
FF2	F	4	pos	3.7	D, 12 mo
FF3	F	10	neg	2	D, 5 mo
FF4	F	14	neg	0.5	A, 6 y
FF5	M	14	neg	2.6	A, 6 y
FF6	M	14	pos	13.7	D, 1 mo
FF7	F	9	pos	6.7	D, 2 mo
FF8	M	1	pos	10.6	D, 7 mo
FF9	F	11	pos	4	D, 9 mo
FF10	F	13	pos	6.7	D, 14 mo
FF11	M	2	neg	4.6	A, 5 y
FF12	F	16	pos	6.5	D, 2 mo
FF13	M	16	pos	8.1	D, 36 mo
FF14	M	13	pos	4.5	D, 11 mo

NOTE: The diagnosis of pGBM was made according to the WHO criteria. Specific histologic findings included regions of necrosis (pseudopalissading necrosis), hypertrophied blood vessels (increased angiogenesis), cells with nuclei highly variable in size and shape with an increased proliferation index. Ras pathway activation was investigated by immunohistochemical staining for FFPE samples or by Western blot analysis for FF samples for pErk. YB-1 expression was assessed by qRT-PCR (FF samples) or by RT-PCR and IHC (FFPE samples).
Abbreviations: IHC, immunohistochemical staining; WB, Western blot analysis; pos, positive; neg, negative; A, alive; D, died.

processed as described previously (27). For both types of samples, slides were scraped if the entire section was diagnosed as GBM, otherwise laser capture microdissection was done to extract pure tumor cell populations. Scraped samples were suspended in proteinase K and incubated overnight at 50°C. Laser capture microdissection, when needed, was done on a PixCell II system (Arcturus Engineering) and 2,000 to 2,500 GBM cells were captured on CapSure HS Caps and processed as per manufacturer's instructions (Arcturus).

RNA extraction and amplification from FFPE and FF tissues; determination of RNA quality. RNA was extracted from FFPE samples using the Paradise reagent system RNA extraction/isolation kit and then amplified using the Paradise reagent system RNA amplification kit (Arcturus). RNA was extracted from FF samples using the PicoPure kit and then amplified using the RiboAmp kit (Arcturus). Fidelity of linear amplified RNA (aRNA) has been previously shown (29); we also validated our material as previously described (30, 31). After the first strand synthesis of the first round, the integrity of the starting material was assessed by PCR reaction with β-actin primers. For each sample, the integrity of aRNA was analyzed on an Agilent 2100 Bioanalyzer (Agilent Technologies) using the RNA 6000 Picochip assay. RNA yield was also measured using Ribogreen RNA quantitation kit (Molecular Probes) on an LS50 luminescence spectrometer (Perkin-Elmer), and the size of the amplification products checked by loading 2 μL on a formaldehyde agarose gel.

RT-PCR. RT-PCR was done in two steps using MMLV-RT (Invitrogen) and random primers. Primers sequences were designed in the 3′ end of each gene with the help of Primer3 software. Primer sequences used for β-actin, PDGFRβ, and Y-Box protein 1 (YB-1) were as follows: β-actin, ATCCCCCAAAGTTCACAATG (forward primer) and GGCTTTTAGGATGGCAAG (reverse primer); PDGFRβ, ATTGCAGGTTGGCACC TTA (forward primer) and TGAGTGAGAAGCACCAGGTTT (reverse primer);YB-1, GCCTGGTTTTTCTCAATACGC (forward primer) and ACAGGTGCTTGCAGTTT GTTG (reverse primer).

Microarray hybridization. For FF samples, 3 to 5 μg of aRNA was converted to cDNA and compared with the same reference pool of cDNA from aRNA extracted from pediatric CB. For FFPE samples, cDNA from 10 μg of aRNA extracted from pGBM samples was compared with the same amount from another reference pool of aRNA extracted from three other pediatric CB. Cy3-labeled or Cy5-labeled cDNA probes from samples and pooled controls were hybridized to human 19K cDNA spotted arrays (19,008 human genes and ESTs; University Health Network). Slides were scanned and fluorescence intensities were quantified using the QuantArray software package (Perkin-Elmer). Inversion of fluors in distinct cDNA probes were done (dye swap) to account for nonspecific dye-associated effects on hybridization and signal detection. We then applied the Lowess scatter smoothing algorithm from the GeneSpring 7.0 software package (Agilent Technologies) to normalize the raw fluorescence data. Sixty-two hybridizations, consisting of dye-swap hybridizations of biological replicates for 31 pediatric samples, were analyzed; analysis of several samples was duplicated to assess for reproducibility from two different RNA extractions and amplifications. The "Filter on Confidence" and ANOVA (Welch *t* test) statistical tools from GeneSpring were used to identify genes with reproducible changes in transcript abundance. In both cases, we applied the Benjamini and Hochberg False Discovery Rate multiple testing correction algorithm. The same software package was used to perform hierarchical clustering and principal component analysis (PCA).

Immunohistochemical analysis. Immunohistochemical analyses for phosphorylated Erk (pErk), glial fibrillary acidic protein (astrocytic marker), and YB-1 were done, and the slides were scored as previously described (32, 33). Slides were counterstained with hematoxylin and mounted. Negative controls (IgG) were included with each batch of sections to confirm the consistency of the analysis. Glial fibrillary acidic protein, a histologically verifiable internal positive control antigen, was used to identify cases in which a lack of immunoreactivity for pErk might indicate problems linked to the labeling of the tissue and thus tissue preservation, rather than a lack of protein phosphorylation. All assays were carried out at the same time with the same reagents. The neuropathologist, blinded to outcome and histology, evaluated the degree of staining.

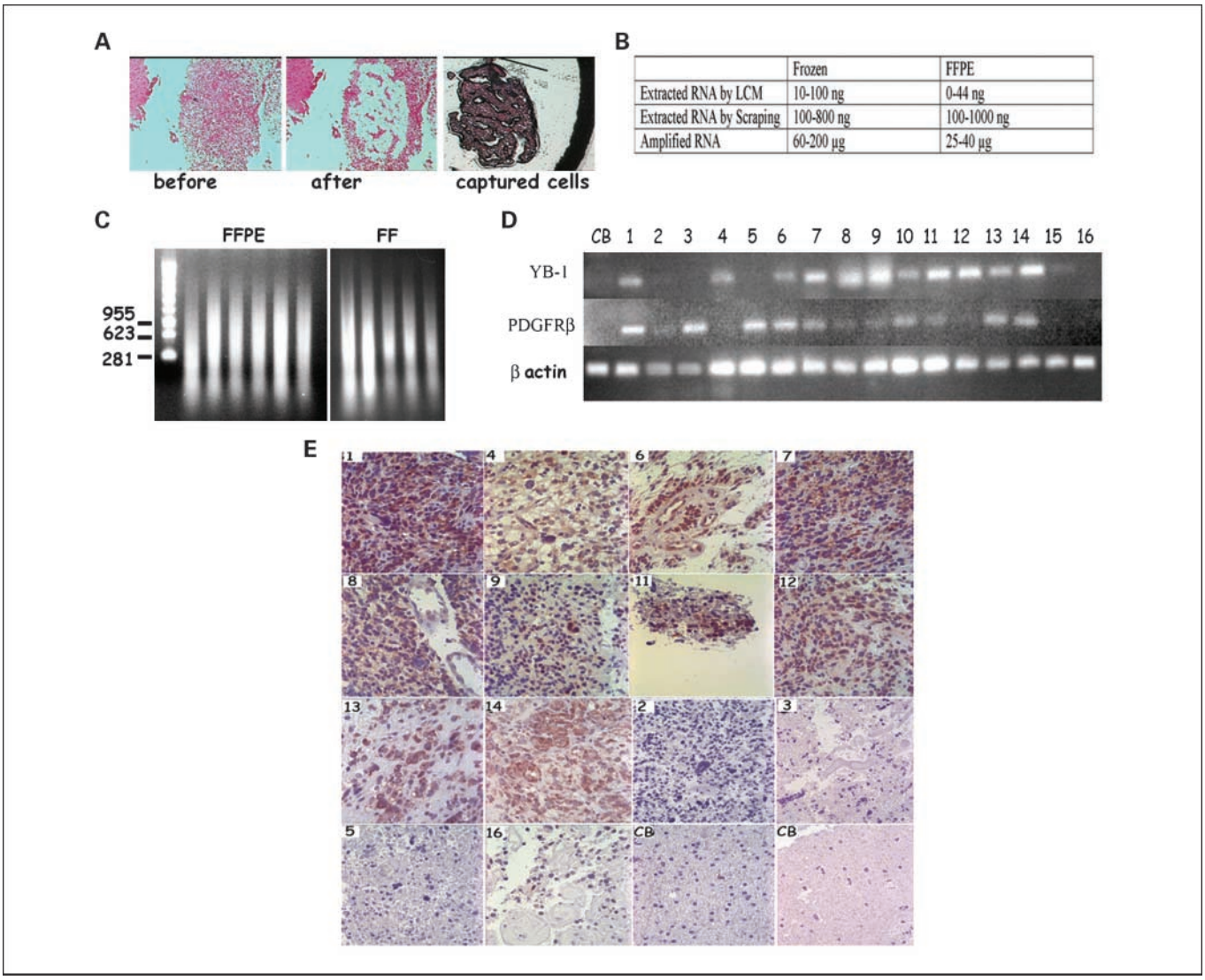
Results

Quantity and quality of aRNA obtained from FFPE samples compared with FF samples. RNA from FF or FFPE samples was extracted after scraping sections or after cell capture using laser capture microdissection (Fig. 1A). RNA yield after extraction was higher after scrape, as expected based on the number of cells (Fig. 1B). To determine if the amount of starting material affects the yield, we did amplifications using a range of 5 to 30 ng of starting RNA. For FFPE samples, total yield of aRNA ranged from 20 to 40 μg after two rounds of amplification and was not affected by the amount of starting RNA material. For FF samples, the yield ranged from 60 to 200 μg and increased with an increase in the amount of starting RNA material (Fig. 1B). aRNA from both sample types showed a similar pattern after electrophoretic separation (Fig. 1C).

To assess the quality of aRNA extracted from either material source, we used the Agilent Bioanalyzer 2100 Picochip. Sharp 18S and 28S rRNA peaks were obtained from FF tissues, and only 18S ribosomal peaks were obtained from FFPE tissues in combination with a characteristic change in the RNA profile indicative of degradation (data not shown). We still did RT-PCR assays using primers designed for the 3′ end of genes that included a housekeeping gene, β-actin, and genes we had previously shown to be overexpressed in a subset of frozen pGBM, *PDGFRβ*, and *YB-1* (31). RT-PCR was done on RNA after extraction and after the first and the second amplification steps. *β-Actin* was present in all samples, whereas *PDGFRβ* transcript was seen in some samples at all steps (Fig. 1D). Data correlated with previous findings obtained from FF pGBM wherein *PDGFRβ* was only up-regulated in a number of samples. *YB-1* was also expressed in some samples (Fig. 1D). Importantly, results for *YB-1* correlated with the immunohistochemical analysis done in parallel on slides from the same FFPE samples, with samples negative by RT-PCR showing no staining for YB-1 and positive samples showing increased protein expression by immunohistochemistry (Fig. 1E).

These results indicate that RNA yield, as well as the quality of extracted RNA, are lower from FFPE samples. However, as previously shown, even if RNA from FFPE samples is degraded, RNA analysis can still be done on this material source.

Gene expression profiling from FFPE samples. The 28S/18S ratio, as measured with the Bioagilent picochip, may provide misleading categorization before microarray analysis (34). Based on the rarity of FF samples from pGBM, we chose to perform microarray analysis on aRNA extracted from FFPE pGBM samples to further validate on an independent data set of pGBM gene expression profiles we obtained on 14 FF pGBM samples (27). When hybridizing human 19K cDNA spotted arrays with 5 μg of cDNA, few specific spots were seen, in marked contrast to FF samples, wherein the number of high-intensity spots was always satisfactory for high-quality hybridizations. We therefore did hybridizations using a range of 10, 15, and 20 μg of cDNA for the same pGBM sample and control. With higher cDNA concentrations (15 or 20 μg), more background noise was obtained, whereas the number of spots



	Frozen	FFPE
Extracted RNA by LCM	10-100 ng	0-44 ng
Extracted RNA by Scraping	100-800 ng	100-1000 ng
Amplified RNA	60-200 μg	25-40 μg

Fig. 1. *A,* laser capture microdissection. FFPE pGBM samples were processed for H&E staining to identify tumor cells. Representative images from a pGBM were taken before and after capture of tumor cells from the same section, and captured cells are shown. *B,* RNA yield before and after linear amplification in FF and FFPE samples. RNA was extracted from slides after scrape or laser capture microdissection, and 5 to 10 ng were subjected to two rounds of T7 RNA polymerase amplification. RNA yield was higher after scrape, and the coefficient of amplification was constantly higher in FF compared with FFPE samples. *C,* Ectrophoretic separation of aRNA from FF and FFPE samples. aRNA (2 μg) from a representative number of samples from FF and FFPE pGBM samples was subjected to electrophoresis to assess for the presence and the size of the amplified material. The size of aRNA was between 300 and 600 bp and similar in both material sources. *D,* RT-PCR detection of β-actin, YB-1, and PDGFRβ gene on RNA extracted from parraffin samples. RT-PCR were done using random primers and starting with 200 ng of aRNA. A reference gene (β-actin) was coamplified with PDGFRβ and YB-1 using gene-specific primers designed within the last 300 bases of each gene. All 16 pGBM FFPE samples and a CB are shown. *β-Actin* is expressed in all samples as expected, whereas *PDGFRβ* and YB-1 expression are detected only in some samples. *E,* pattern of YB-1 expression in FFPE pGBM. Immunohistochemical staining for YB-1 was done on 14 pGBM and 2 CBs. Anti – YB-1 COOH terminus antibody staining and scoring of slides were done as previously described (47). Staining confirms results obtained by RT-PCR on the same samples.

showing specific hybridization on the slide did not increase (data not shown). Optimal results and reproducible results on 13 of 16 FFPE pGBM samples were obtained for 10 μg of amplified cDNA even if, when compared with aRNA from FF samples, the background noise was higher and a significantly lower number of spots were seen on the array slides.

GeneSpring's Filter on Confidence tool identified a list of transcripts with statistically significant changes in abundance in FFPE pGBM samples compared with the pooled controls (Welch t test, $P < 0.05$; multiple testing correction: Benjamini and Hochsberg, false discovery rate of 3.4%). Other statistical algorithms (the Wilcoxon-Mann-Whitney test of significance analysis of microarrays) were tested with similar results. Differentially expressed transcripts between pGBM and CB were obtained for FF or FFPE pGBM samples. Analysis of the data showed 777 transcripts to be differentially expressed in FFPE samples compared with the pooled control and 3,647 modulated transcripts in FF samples compared with the FF-pooled CB. A two-dimensional graph and two-dimensional hierarchical clustering organized and visualized the profiles of these differentially regulated transcripts (Y axis) from each of the 13 FFPE pGBM samples and the previously reported 14 FF pGBM samples (X axis; Fig. 2B and C, respectively; ref. 27). Both the graph representing differentially expressed transcripts between pGBM samples and the pooled controls and the cluster tree of these differentially expressed transcripts show that a

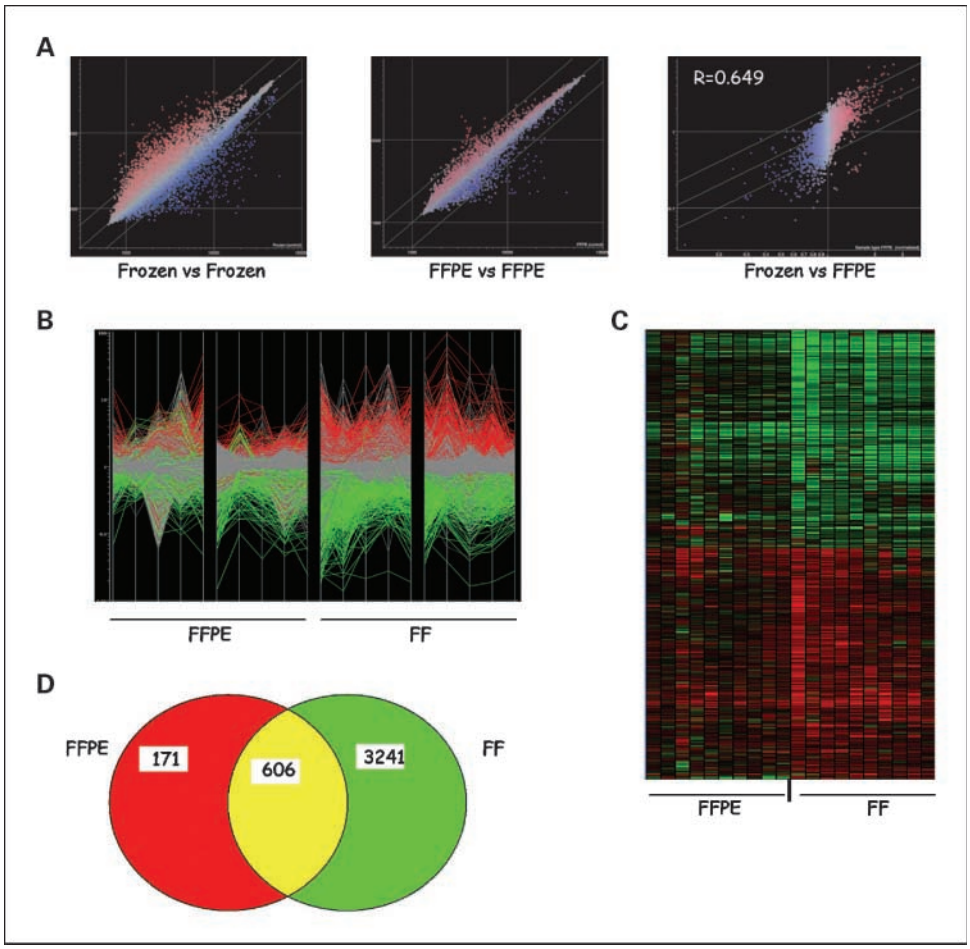


Fig. 2. Gene expression profiling of FFPE samples reproduces the pattern of expression profiles obtained on FF pGBM samples. *A,* profile similarities between different subgroups of GBM. Scatter plots comparing the change in transcript abundance between pairs of subgroups of pGBM: FF pGBM (*left*), FFPE (*middle*), and FFPE versus FF (*right*). Transcripts colored in red (up-regulated) and blue (down-regulated) show a statistical change in abundance between the pair of subgroups of GBM analyzed relative to the control (Welsh *t* test, $P_{cutoff} < 0.05$; multiple testing correction: Benjamini and Hochsberg False Discovery Rate). Log^2 intensity scatter plots were generated using raw intensity data, and Pearson correlation coefficients were calculated for FF versus FFPE samples (*left*; $r = 0.649$). *B,* unsupervised hierarchical clustering of the probes with a statistically significant change in transcript abundance (Welsch *t* test, $P < 0.05$; Benjamini and Hochsberg) between FF pGBM, FFPE pGBM, and the pool of normal brain tissue (*X* axis) shows a decrease in the number and the fold change of transcript profiles in FFPE samples (probes on the *Y* axis). Each experimental data point is colored according to the change in fluorescence ratio: more abundant in pGBM (*red*) and less abundant colored (*green*). *C,* two-dimensional hierarchical clustering of 486 probes that exhibit the highest statistically significant change in transcript abundance between sample pairs in FF samples (Welsch *t* test, $P < 0.05$; Benjamini and Hochsberg; see Supplementary gene lists) shows that the pattern of overexpressed/down-regulated genes is similar in FFPE samples. Each experimental data point is colored according to the change in fluorescence ratio: more abundant in pGBM (*red*) and less abundant colored (*green*). *D,* Venn diagram of differentially regulated transcripts relative to the pooled CB in FFPE pGBM samples ($n = 777$) and in FF samples ($n = 3847$; Welch *t* test, $P < 0.05$) showing a significant overlap ($n = 606$) between both lists of transcripts (Supplementary gene lists).

significant number of transcripts are lost in the FFPE compared with the FF samples. Also, as shown on the two-dimensional graph, the clustering tree, and the scatter plot graph of all samples from both material sources, transcripts from the FFPE samples indicate a decreased fold change relative to the CB when compared with the same transcripts from FF samples (Fig. 2A). Despite these differences in number of differentially expressed transcripts and fold change and the use of different controls, we still found a reasonable correlation of $R^2 = 0.649$ between both sample types. Importantly, a similar pattern of gene expression profiles was maintained between FF and FFPE pGBM samples: transcripts that were overexpressed (color coded in red) in the FF samples were also overexpressed in the FFPE, and conversely, transcripts that were down-regulated in both sample types were similar (color coded in green; Fig. 2B and C). This indicates that, even if we are losing a number of differentially regulated transcripts in the FFPE samples compared with the FF samples, as well as the extent of the fold change, we are maintaining a similar expression pattern for the significant differentially regulated transcripts in FFPE samples.

Analysis of gene expression profiles from FFPE samples mirrors results obtained from FF samples. We next used Venn diagrams to compare the list of significantly modulated transcripts for both sample sets we generated using the *P*-value cutoff of 0.05. This cutoff was chosen as many transcripts on the verge of statistical significance were missed by a more stringent *P* value, as also previously described in studies comparing different platforms or RNA from different material sources with variable sensitivity for detecting transcripts, including weakly expressed genes (34–38). We saw an overlap between both gene lists of 606 genes shared by the two subsets of pGBM samples, which gives a significance of

overlap of 1.0e-09 according to the Fischer's exact test (Fig. 2D; Supplementary gene lists). We analyzed the data set using a module-level view obtained from a "cancer compendium" (39) and also organized the gene sets using GoMiner, a computer resource that incorporates the hierarchical structure of the gene ontology consortium (2) to automate a functional categorization of gene lists based on biological processes. Both methods aim to distill a higher order from a large list of genes. They yielded similar results, showing that modules and gene ontology terms overlapped for the top 15 categories obtained on significantly modulated transcripts from FF and FFPE data sets (Table 2). The use of different CB could partly account for the lack of complete overlap between both types of samples on gene lists. We had previously determined for FF samples that different CB had similar, if not totally overlapping, gene expression profiles (Pearson correlation, $r = 0.93$). When categorizing the nature of these 171 genes by gene ontology processes, they were mainly grouped into catabolic transcripts (Supplementary gene lists). When we looked at the statistically significant transcripts present in both gene lists, most of the genes that we had validated by qRT-PCR in FF samples and chose to pursue based on their potential involvement in the oncogenesis of pGBM were also present when analyzing significantly modulated transcripts in FFPE samples (Table 3; Supplementary gene lists).

We have previously shown using FF samples that there are at least two prognostic subgroups of pGBM that could be molecularly separated based on their association or not with evidence consistent with an aberrant Ras active pathway. The presence of phosphorylated effectors of Ras, including pErk1/pErk2, phosphorylated MAP/ERK kinase 1/MAP/ERK kinase 2, and phosphorylated Raf, investigated by Western in FF samples and a positive pErk1/pErk2 staining by immunohistochemistry in FFPE samples, were considered consistent with Ras activation (Fig. 3A; refs. 27, 32). To further validate these findings, the same analysis was done on FFPE samples using PCA, a method of data reduction in which the high dimensionality of the data is reduced to two to three viewable dimensions representing linear combinations of genes that account for most of the variance of the data set and which allows to visualize similarities within samples. Similarly to our findings in FF samples (27), PCA separated both sets of FFPE and FF pGBM samples into two groups, indicating the presence of at least two distinct populations of pGBM (Fig. 3B). The two pGBM populations segregated by PCA were associated with evidence consistent with differing Ras activity in samples (Fig. 3B). Also, ANOVA testing was able to identify transcripts that could distinguish tumors associated with differing Ras activity in both FF and FFPE samples while showing a significant degree of correlation between frozen and FFPE (linear correlation in the scatter plots; Fig. 3C).

Table 2. Gene ontology classification using gene oncology miner of the differentially expressed transcripts relative to the pooled CBs of FF and FFPE pGBM samples

	FFPE	FF
1.	Physiologic process	Cellular process
2.	Cellular process	Physiologic process
3.	Cellular physiologic process	Cellular physiologic process
4.	Metabolism	Cell communication
5.	Cellular metabolism	Macromolecule metabolism
6.	Primary metabolism	Regulation of biological processes
7.	Macromolecule metabolism	Cellular macromolecule metabolism
8.	Cellular macromolecule metabolism	Protein metabolism
9.	Protein metabolism	Cellular protein metabolism
10.	Cellular protein metabolism	Regulation of cellular process
11.	Localization	Regulation of physiologic processes
12.	Establishment of localization	Localization
13.	Organismal physiologic process	Establishment of localization
14.	Transport	Regulation of cellular physiologic processes
15.	Regulation of biological process	Biopolymer metabolism

NOTE: This classification shows that the top 15 biological processes affected in pGBM are similar in both sample types.

Discussion

Validation of data obtained from gene expression profiling on independent data sets is crucial. We show herein that RNA from FFPE samples can be successfully isolated from laser-captured or scraped samples amplified and used for reliable single-transcript or multiple-transcript analysis. We did several RT-PCR analyses on RNA extracted from the material source for housekeeping and tumor-modulated transcripts and validated results by immunohistochemical analysis for the 19 FFPE samples included in this study. Importantly, using cDNA arrays, we were able to perform gene expression profiling on this type of material for 13 of 16 pGBM and 3 of 3 CB samples (84%). Results of expression profiles helped validate the data sets and observations we had previously generated on FF pGBM, even if the number of differentially regulated transcripts and the degree of fold change in transcripts were decreased in FFPE samples.

Several groups showed that the extraction of RNA from fixed samples is possible even if only small targets were being amplified, with an average of 300 bp in length of RNA extracted from fixed samples (5, 7, 9, 11–13, 40). These groups further showed the feasibility of single or multiplex transcript analysis, including RT-PCR and qRT-PCR, on a limited number of RNA transcripts extracted from fixed tissues. In this study, we were able to extract sufficient quality and quantity of RNA material for gene expression profiling. RNA from FFPE samples produced higher background, and the number of significantly modulated transcripts was decreased, as also previously shown in another study that used cDNA-mediated annealing, selection, extension, and ligation (41). However, we were still able to reproduce findings we first observed on FF samples. Using PCA, we were able to identify two subsets of pGBM in FFPE pGBM samples based on their association or not with evidence consistent with an aberrantly active Ras pathway, similar to findings on FF pGBM samples. Analysis of the data set using

Table 3. List of the top 100 differentially expressed transcripts relative to the pooled CBs common to FF and FFPE pGBM

Common name	Description
sparc	Secreted protein, acidic, cysteine-rich (osteonectin)
rbp1	Retinol binding protein 1, cellular
hla-dpa1	Major histocompatibility complex, class II, DPα1
cfl1	Cofilin 1 (nonmuscle)
hla-drb3	Major histocompatibility complex, class II, DRβ4
col6a1	Collagen, type VI, α1
rpl10	Ribosomal protein L10
hcn3	Hyperpolarization-activated cyclic nucleotide-gated potassium channel 3
arhc	Ras homologue gene family, member C
api5	Apoptosis inhibitor 5
timp1	Tissue inhibitor of metalloproteinase 1 (erythroid potentiating activity)
hla-dpb1	Major histocompatibility complex, class II, DPβ1
s100a6	S100 calcium binding protein A6 (calcyclin)
tcfl1	Transcription factor-like 1
nsep1	Nuclease-sensitive element-binding protein 1 (Y-Box protein 1)
serf2	Small EDRK-rich factor 2
cd63	CD63 antigen (melanoma 1 antigen)
chi3l1	Chitinase 3-like 1 (cartilage glycoprotein-39)
myl6	Myosin, light polypeptide 6, alkali, smooth muscle and nonmuscle
arha	Ras homologue gene family, member A
gapd	Glyceraldehyde-3-phosphate dehydrogenase
cd99	CD99 antigen
ifi30	Interferon, γ-inducible protein 30
fn1	Fibronectin 1
rps3	Ribosomal protein S3
tcf8	Transcription factor 8 (represses interleukin 2 expression)
apoc1	Apolipoprotein C-I
rpl10	Ribosomal protein L10
rpl13a	Ribosomal protein L13a
rpl41	Ribosomal protein L41
calu	Calumenin
gns	Glucosamine (*N*-acetyl)-6-sulfatase (Sanfilippo disease IIID)
rps16	Ribosomal protein S16
rps17	Ribosomal protein S17
hla-dpb1	Major histocompatibility complex, class II, DPβ1
pabpc4	Poly(A) binding protein, cytoplasmic 4 (inducible form)
cdk4	Cyclin-dependent kinase 4
tm4sf1	Transmembrane 4 superfamily member 1
cd63	CD63 antigen (melanoma 1 antigen)
polr2b	Insulin-like growth factor binding protein 7
sox10	SRY (sex determining region Y)-box 10
rpl13a	Ribosomal protein L13a
eif3s4	Eukaryotic translation initiation factor 3, subunit 4 δ, 44 kDa
nrg3	Neuregulin 3
pltp	Phospholipid transfer protein
igkc	HRV Fab N8-VL
h3f3a	H3 histone, family 3B (H3.3B)
mdk	Midkine (neurite growth-promoting factor 2)
rps3	Ribosomal protein S3

Table 3. List of the top 100 differentially expressed transcripts relative to the pooled CBs common to FF and FFPE pGBM (Cont'd)

Common name	Description
col4a2	Collagen, type IV, α2
h2afz	H2A histone family, member Z
ighm	Immunoglobulin λ constant 2 (Kern-Oz marker)
arhc	Ras homologue gene family, member C
nnmt	Nicotinamide *N*-methyltransferase
ighm	Hepatitis B surface antigen antibody variable domain
rpl8	Ribosomal protein L8
eef1g	Eukaryotic translation elongation factor 1γ
snrpb	Small nuclear ribonucleoprotein polypeptides B and B1
mect1	Mucoepidermoid carcinoma translocated 1
ncl	Nucleolin
mage-e1	Melanoma antigen family D, 4
xpo1	Exportin 1 (CRM1 homologue, yeast)
ifitm1	Interferon-induced transmembrane protein 1 (9–27)
dad1	Defender against cell death 1
plat	Plasminogen activator, tissue
ranbp1	RAN binding protein 1
nme1	Nonmetastatic cells 1, protein (NM23A) expressed in
rplp2	Ribosomal protein, large P2
ccnd1	Cyclin D1 (PRAD1: parathyroid adenomatosis 1)
pp1201	H19, imprinted maternally expressed untranslated mRNA
myl6	Myosin, light polypeptide 6, alkali, smooth muscle and nonmuscle
igkc	Immunoglobulin κ constant
fabp7	Fatty acid binding protein 7, brain
canx	Calnexin
pgd	Phosphogluconate dehydrogenase
gapd	Glyceraldehyde-3-phosphate dehydrogenase
s100a10	S100 calcium binding protein A10 [annexin II ligand, calpactin I, light polypeptide (p11)]
timp1	Tissue inhibitor of metalloproteinase 1 (erythroid potentiating activity, collagenase inhibitor)
hmgn1	High-mobility group nucleosome binding domain 1
psmb2	Proteasome (prosome, macropain) subunit, β type, 2
hbb	Hemoglobin, β
rpl27	Ribosomal protein L27
col4a2	Collagen, type IV, α2
pkm2	Pyruvate kinase, muscle
aebp1	AE binding protein 1
actn4	Actinin, α4
pml	Promyelocytic leukemia
rpl18a	Thioredoxin reductase 1
tubb	Tubulin, β polypeptide paralogue
tubb2	Tubulin, β, 2
dgkd	Diacylglycerol kinase, δ 130 kDa
nedd5	Septin 2
rpn2	Ribophorin II
ercc1	Excision repair cross-complementing rodent repair deficiency, complementation group 1
slc20a1	Solute carrier family 20 (phosphate transporter), member 1
rpl18	Ribosomal protein L18
rps19	Ribosomal protein S19
dusp6	Dual specificity phosphatase 6
polr2l	Polymerase (RNA) II (DNA directed) polypeptide L, 7.6 kDa

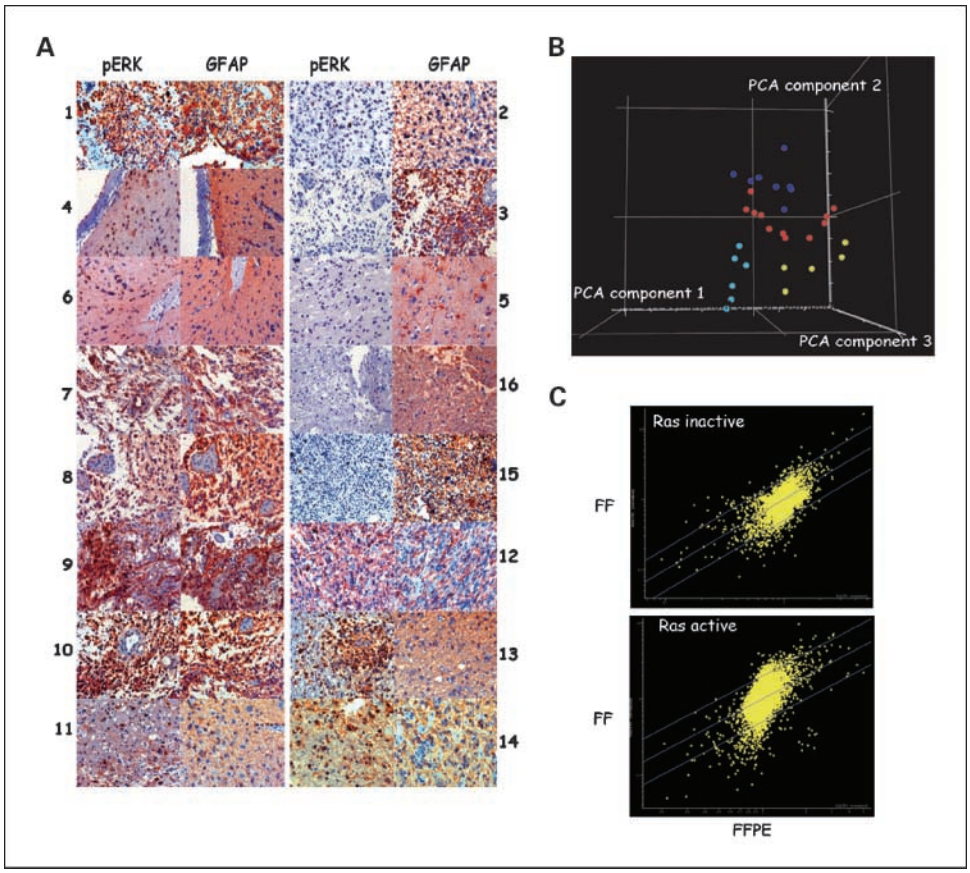


Fig. 3. Tumor samples show distinct expression profiles that correlate with Ras pathway activation. *A,* immunohistochemical analyses for pErk, glial fibrillary acidic protein (astrocytic marker) were done for the 16 pGBM samples included in this study. Samples are numbered as in Table 1 and Fig. 1D. *B,* the 13 FFPE and 14 FF pGBM samples were subjected to a PCA based on the expression profile measured on 15,068 individual probes. A three-dimensional plot of PCA components 1, 2, and 3 distinguished the Ras scores of the pediatric tumors irrespective of the nature of the sample source. Samples are color-coded for clarity issues: pGBM samples associated with active Ras pathway (*red*, FFPE; *dark blue*, FF); pGBM samples not associated with active Ras (*yellow*, FFPE; *light blue*, FF). Three FFPE samples with an active Ras pathway for which we had enough material were treated in duplicate with separate RNA extraction, amplification, and hybridization. They migrated similarly on the PCA graph, further confirming the reproducibility of the gene expression analysis on FFPE samples. *C,* ANOVA identified differentially expressed transcripts in both sample sources based on their association/lack of association with an active Ras pathway. Log^2 intensity scatter plots were generated using raw intensity data, and Pearson correlation coefficients calculated for FF versus FFPE samples associated with Ras activation ($r = 0.651$) and pGBM samples not associated with Ras activation ($r = 0.712$).

GoMiner showed an overlap in biological processes affected in pGBM in both sample types (Table 2). The list of differentially regulated transcripts in pGBM compared with the CBs extracted from the FFPE data sets (Supplementary gene lists) was concordant with the one obtained in FF samples and included transcripts that were previously identified by other profiling studies on GBM (42, 43). Lack of total overlap between the transcripts from the gene lists from both sample types may be due to the use of different CB. Alternately, the use of material from different sources with different degrees of tissue preservation and different fixation procedures and durations can introduce a bias as suggested by other studies (18, 44). Indeed, fixatives like ethanol or Bouin are considered to be less detrimental to RNA integrity and preserve tissue architecture and cell morphology. However, in most of these studies, the analysis was done on samples manipulated *de novo* under conditions optimized for the best results. This is not true for most archival tissues, which are mainly formalin-fixed through the choice of the pathology departments and handled without consideration of future RNA extraction. This makes the type of fixative and different handling procedures unavoidable sources of variability when using archival materials. However, as shown by our results, this variability does not prevent reliable microarray analysis.

In a recently published study, Penland et al. profiled 157 archival FFPE colon samples and show similar results to this study in terms of RNA yield, size, and degradation profile (45). However, only 24% of these FFPE samples yielded reliable microarray data. The authors went to establish a stringent algorithm to obtain a more precise indication of the extent of RNA degradation in a sample before hybridization on slides. This included a Taqman assay designed to determine 3′-to-5′ ratios in total RNA extracted from a sample, which increased their success rate from 17% to 48% and the control for efficiency of labeling. In this study, we had a significant success rate in this regard (84%). This may be due to technical factors and/or to the autolytic properties of the tissue itself (brain as opposed to colon). One could also consider a bias toward a smaller sample size as in another recent study of five FFPE sample colon cancers that used the same RNA extraction, amplification, and array platform than Penland et al.; the success rate for microarray analysis was 100% (46). In all, for larger scale studies wherein the rate limiting aspect is the efficiency of hybridizations and not the small number of samples available, a more thorough screening should be done to optimize the success rate of the hybridization and limit undue work and costs (45).

In summary, FF samples are the obvious choice for optimal transcript analysis. However, archival fixed tissues are an invaluable resource. Despite the numerous challenges faced by their utilization for microarrays, our data suggest that, in rare tumors and in cases wherein the lack of clinical data is an issue, the use of these archival FFPE samples to investigate expression profiles is feasible. In this instance, both the availability of a larger scale of samples and its association to clinical data help compensate for lower RNA quality, providing researchers with a valuable source of material that can be used for validation and further investigations in rare diseases.

References

1. Segal E, Friedman N, Kaminski N, Regev A, Koller D. From signatures to models: understanding cancer using microarrays. Nat Genet 2005;37 Suppl:S38–45.
2. Zeeberg BR, Feng W, Wang G, et al. GoMiner: a resource for biological interpretation of genomic and proteomic data. Genome Biol 2003;4:R28.
3. Lewis F, Maughan NJ, Smith V, Hillan K, Quirke P. Unlocking the archive-gene expression in paraffin-embedded tissue. J Pathol 2001;195:66–71.
4. Masuda N, Ohnishi T, Kawamoto S, Monden M, Okubo K. Analysis of chemical modification of RNA from formalin-fixed samples and optimization of molecular biology applications for such samples. Nucleic Acids Res 1999;27:4436–43.
5. Jackson DP, Lewis FA, Taylor GR, Boylston AW, Quirke P. Tissue extraction of DNA and RNA and analysis by the polymerase chain reaction. J Clin Pathol 1990;43:499–504.
6. KorblerT, Grskovic M, Dominis M, Antica M. A simple method for RNA isolation from formalin-fixed and paraffin-embedded lymphatic tissues. Exp Mol Pathol 2003;74:336–40.
7. Specht K, Richter T, Muller U, Walch A, Werner M, Hofler H. Quantitative gene expression analysis in microdissected archival formalin-fixed and paraffin-embedded tumor tissue. Am J Pathol 2001;158: 419–29.
8. Walch A, Specht K, Smida J, et al. Tissue microdissection techniques in quantitative genome and gene expression analyses. Histochem Cell Biol 2001;115: 269–76.
9. Godfrey TE, Kim SH, Chavira M, et al. Quantitative mRNA expression analysis from formalin-fixed, paraffin-embedded tissues using 5′ nuclease quantitative reverse transcription-polymerase chain reaction. J Mol Diagn 2000;2:84–91.
10. Finke J, Fritzen R, Ternes P, Lange W, Dolken G. An improved strategy and a useful housekeeping gene for RNA analysis from formalin-fixed, paraffin-embedded tissues by PCR. Biotechniques 1993;14:448–53.
11. Stanta G, Bonin S, Perin R. RNA extraction from formalin-fixed and paraffin-embedded tissues. Methods Mol Biol 1998;86:23–6.
12. Stanta G, Bonin S, Utrera R. RNA quantitative analysis from fixed and paraffin-embedded tissues. Methods Mol Biol 1998;86:113–9.
13. Steg A, Wang W, Blanquicett C, et al. Multiple gene expression analyses in paraffin-embedded tissues by TaqMan low-density array: Application to hedgehog and Wnt pathway analysis in ovarian endometrioid adenocarcinoma. J Mol Diagn 2006;8:76–83.
14. Capodieci P, Donovan M, Buchinsky H, et al. Gene expression profiling in single cells within tissue. Nat Methods 2005;2:663–5.
15. Ma XJ, Patel R, Wang X, et al. Molecular classification of human cancers using a 92-gene real-time quantitative polymerase chain reaction assay. Arch Pathol Lab Med 2006;130:465–73.
16. Pagedar NA, Wang W, Chen DH, et al. Gene expression analysis of distinct populations of cells isolated from mouse and human inner ear FFPE tissue using laser capture microdissection-a technical report based on preliminary findings. Brain Res 2006;1091: 289–99.
17. Tothill RW, Kowalczyk A, Rischin D, et al. An expression-based site of origin diagnostic method designed for clinical application to cancer of unknown origin. Cancer Res 2005;65:4031–40.
18. Kabbarah O, Pinto K, Mutch DG, Goodfellow PJ. Expression profiling of mouse endometrial cancers microdissected from ethanol-fixed, paraffin-embedded tissues. Am J Pathol 2003;162:755–62.
19. Maher EA, Furnari FB, Bachoo RM, et al. Malignant glioma: genetics and biology of a grave matter. Genes Dev 2001;15:1311–33.
20. Holland EC. Gliomagenesis: genetic alterations and mouse models. Nat Rev Genet 2001;2:120–9.
21. Zhu Y, Parada LF. The molecular and genetic basis of neurological tumours. Nat Rev Cancer 2002;2: 616–26.
22. Kleihues P, Louis DN, Scheithauer BW, et al. The WHO classification of tumors of the nervous system. J Neuropathol Exp Neurol 2002;61:215–25;discussion 226–9.
23. Packer RJ. Primary Central Nervous System Tumors in Children. Curr Treat Options Neurol 1999; 1:395–408.
24. Louis DN, Holland EC, Cairncross JG. Glioma classification: a molecular reappraisal. Am J Pathol 2001; 159:779–86.
25. Pollack IF, Hamilton RL, Finkelstein SD, Lieberman F. Molecular abnormalities and correlations with tumor response and outcome in glioma patients. Neuroimaging Clin N Am 2002;12:627–39.
26. Rood BR, Macdonald TJ. Pediatric high-grade glioma: molecular genetic clues for innovative therapeutic approaches. J Neurooncol 2005.
27. Faury D, Nantel A, Dunn SE, et al. Molecular profiling identifies prognostic subgroups of pediatric glioblastoma and shows increased YB-1 expression in tumors. J Clin Oncol 2007;25:1196–208.
28. Mariani L, McDonough WS, Hoelzinger DB, et al. Identification and validation of P311 as a glioblastoma invasion gene using laser capture microdissection. Cancer Res 2001;61:4190–6.
29. Aoyagi K, Tatsuta T, Nishigaki M, et al. A faithful method for PCR-mediated global mRNA amplification and its integration into microarray analysis on laser-captured cells. Biochem Biophys Res Commun 2003; 300:915–20.
30. Polacek DC, Passerini AG, Shi C, et al. Fidelity and enhanced sensitivity of differential transcription profiles following linear amplification of nanogram amounts of endothelial mRNA. Physiol Genom 2003; 13:147–56.
31. Faury D, Nantel A, Dunn SE, et al. Molecular profiling identifies prognostic subgroups of pediatric glioblastoma. J Clin Oncol 2007;75:1196–208.
32. Kreisberg JI, Malik SN, Prihoda TJ, et al. Phosphorylation of Akt (Ser473) is an excellent predictor of poor clinical outcome in prostate cancer. Cancer Res 2004;64:5232–6.
33. Pollack IF, Finkelstein SD, Woods J, et al. Expression of p53 and prognosis in children with malignant gliomas. N Engl J Med 2002;346:420–7.
34. Copois V, Bibeau F, Bascoul-Mollevi C, et al. Impact of RNA degradation on gene expression profiles: Assessment of different methods to reliably determine RNA quality. J Biotechnol 2006.
35. Shi L, Reid LH, Jones WD, et al. The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. Nat Biotechnol 2006;24:1151–61.
36. Patterson TA, Lobenhofer EK, Fulmer-Smentek SB, et al. Performance comparison of one-color and two-color platforms within the MicroArray Quality Control (MAQC) project. Nat Biotechnol 2006;24:1140–50.
37. Canales RD, Luo Y, Willey JC, et al. Evaluation of DNA microarray results with quantitative gene expression platforms. Nat Biotechnol 2006;24:1115–22.
38. Minor JM. Microarray quality control. Methods Enzymol 2006;411:233–55.
39. Segal E, Friedman N, Koller D, Regev A. A module map showing conditional activity of expression modules in cancer. Nat Genet 2004;36:1090–8.
40. Gloghini A, Canal B, Klein U, et al. RT-PCR analysis of RNA extracted from Bouin-fixed and paraffin-embedded lymphoid tissues. J Mol Diagn 2004;6: 290–6.
41. Bibikova M, Talantov D, Chudin E, et al. Quantitative gene expression profiling in formalin-fixed, paraffin-embedded tissues using universal bead arrays. Am J Pathol 2004;165:1799–807.
42. Rich JN, Hans C, Jones B, et al. Gene expression profiling and genetic markers in glioblastoma survival. Cancer Res 2005;65:4051–8.
43. Shi Q, Bao S, Maxwell JA, et al. Secreted protein acidic, rich in cysteine (SPARC), mediates cellular survival of gliomas through AKT activation. J Biol Chem 2004;279:52200–9.
44. Karsten SL, Van Deerlin VM, Sabatti C, Gill LH, Geschwind DH. An evaluation of tyramide signal amplification and archived fixed and frozen tissue in microarray gene expression analysis. Nucleic Acids Res 2002;30:E4.
45. Penland SK, Keku TO, Torrice C, et al. RNA expression analysis of formalin-fixed paraffin-embedded tumors. Lab Invest 2007;87:383–91.
46. Coudry RA, Meireles SI, Stoyanova R, et al. Successful application of microarray technology to microdissected formalin-fixed, paraffin-embedded tissue. J Mol Diagn 2007;9:70–9.
47. Sutherland BW, Kucab J, Wu J, et al. Akt phosphorylates the Y-box binding protein 1 at Ser102 located in the cold shock domain and affects the anchorage-independent growth of breast cancer cells. Oncogene 2005;24:4281–92.