

Methodology article

## Obtaining reliable information from minute amounts of RNA using cDNA microarrays

Limei Hu<sup>1</sup>, Jing Wang<sup>2</sup>, Keith Baggerly<sup>2</sup>, Hua Wang<sup>1</sup>, Gregory N Fuller<sup>1</sup>, Stanley R Hamilton<sup>1</sup>, Kevin R Coombes<sup>2</sup> and Wei Zhang\*<sup>1</sup>

Address: <sup>1</sup>Cancer Genomics Core Laboratory-Department of Pathology, The University of Texas M D Anderson Cancer Center, 1515 Holcombe Blvd, Houston, TX 77030, USA and <sup>2</sup>Department of Biostatistics, The University of Texas M D Anderson Cancer Center, 1515 Holcombe Blvd, Houston, TX 77030, USA

E-mail: Limei Hu - lhu@mdanderson.org; Jing Wang - jwang@odin.mdacc.tmc.edu; Keith Baggerly - kbaggerly@odin.mdacc.tmc.edu; Hua Wang - huawang@mdanderson.org; Gregory N Fuller - gfuller@mdanderson.org; Stanley R Hamilton - srhamilton@mdanderson.org; Kevin R Coombes - krc@odin.mdacc.tmc.edu; Wei Zhang\* - wzhang@mdanderson.org

\*Corresponding author

Published: 21 June 2002

*BMC Genomics* 2002, **3**:16

This article is available from: <http://www.biomedcentral.com/1471-2164/3/16>

Received: 5 June 2002

Accepted: 21 June 2002

© 2002 Hu et al; licensee BioMed Central Ltd. Verbatim copying and redistribution of this article are permitted in any medium for any purpose, provided this notice is preserved along with the article's original URL.

### Abstract

**Background:** High density cDNA microarray technology provides a powerful tool to survey the activity of thousands of genes in normal and diseased cells, which helps us both to understand the molecular basis of the disease and to identify potential targets for therapeutic intervention. The promise of this technology has been hampered by the large amount of biological material required for the experiments (more than 50 µg of total RNA per array). We have modified an amplification procedure that requires only 1 µg of total RNA. Analyses of the results showed that most genes that were detected as expressed or differentially expressed using the regular protocol were also detected using the amplification protocol. In addition, many genes that were undetected or weakly detected using the regular protocol were clearly detected using the amplification protocol. We have carried out a series of confirmation studies by northern blotting, western blotting, and immunohistochemistry assays.

**Results:** Our results showed that most of the new information revealed by the amplification protocol represents real gene activity in the cells.

**Conclusion:** We have confirmed a powerful and consistent cDNA microarray procedure that can be used to study minute amounts of biological tissue.

### Background

Cancer is a progressive genetic disease that involves the accumulation of multiple and heterogeneous genetic and epigenetic changes. Among the 30,000 – 40,000 human genes [1,2], an expanding list of genes has been shown to be involved in cancer development and progression. Recently developed technologies, including cDNA microar-

rays, allow cancer researchers to screen thousands of genes simultaneously to identify genes that show abnormal expression in cancers [3–6].

Cancer researchers face a serious challenge when trying to apply these tools to clinical cancer specimens. First, cancers are often detected at a late stage in development, after

multiple genetic and epigenetic changes have rendered the cancer metastatic and highly refractory even to harsh treatment. Multiple clones in the same tumor mass, which can often be distinguished by morphological features, would yield more information if they could be microdissected and analyzed separately. Second, tumor tissues invariably include a mixture of different cells, such as inflammatory cells, stromal cells, vascular cells, and others. Each type of cells may contribute to a unique aspect of the cancer phenotype and may serve as a target for treatment. Therefore, there is a need to dissect those cells for separate study. For example, it was found that tumor vascular cells have distinct gene expression profiles from those of normal vascular cells [7], and the genes that are uniquely expressed on the surface of tumor endothelium cells can serve as targets for specific cancer therapy. Finally, a key breakthrough in cancer treatment would come from early detection and diagnosis of cancer and the study of molecular events in the early stages of cancer progression. A difficulty in studying cancer at early stages is that the tumor size is small and only limited material can be obtained through means such as fine needle aspiration. Identification of the genetic changes in such a small sample is especially challenging because only limited assays can be performed using conventional approaches. It is obvious that the full potential of genomic technologies will only be realized if they can be applied to minute amounts of biological material.

For a typical gene expression profiling experiment carried out on a glass microarray where thousands of cDNA probes are deposited in an orderly manner, RNA is first isolated from the biological material under study. The mRNA (1–5% of total RNA) is reverse transcribed to cDNA, during which process fluorescent dyes (Cy3 or Cy5) are incorporated. The labeled cDNAs are hybridized to the microarray and, after washing, the fluorescent signals are detected by a laser scanner. Current protocols typically require more than 50 µg of total RNA for consistent microarray hybridization. Several hundred milligrams of tumor tissues are often needed to obtain this much RNA, which is simply unavailable in many situations. Thus, it is crucial to develop a more sensitive and reliable procedure that requires less RNA. Several molecular biology approaches, such as T7 in vitro transcription and PCR based assays, have been attempted [8–10]. However, only cursory analysis has been carried out to validate the assays, and limited confirmation experiments have been performed to evaluate the validity of the amplification results.

In this study, we performed and analyzed a set of nine microarray experiments to evaluate an amplification protocol adopted from Wang *et al*[9]. Focusing primarily on the ability to identify differentially expressed genes, we

developed the following criteria for a successful amplification protocol:

1. Because amplification may enhance the signal of genes expressed at low copy numbers, more genes should be detected by an amplification protocol than by a regular protocol.
2. Most genes detected as differentially expressed using a regular protocol should also be detected using an amplification protocol. In other words, the two protocols should reveal similar patterns of differential expression.
3. An amplification protocol should generate signal intensity profiles as reproducibly and reliably as a regular protocol.
4. Microarray results obtained from an amplification protocol should match data obtained from other molecular biology approaches such as northern blotting, western blotting and immunohistochemistry assay.

Our results showed that our amplification protocol produced reproducible, reliable microarray data that was consistent with the regular protocol. We also confirmed that our amplification protocol revealed accurate information about the differential expression of low copy number genes that failed to give sufficient signal intensities using the regular protocol. Therefore, many clinical experiments for which only a minute amount of material is available can be pursued using this protocol.

## Results

A total of nine microarray experiments were performed (yielding 18 images). Among the nine experiments, five used the regular protocol with varying amounts of total RNA (100 µg to 300 µg) and four used the amplification protocol with 1.0 µg total RNA. Each experiment compared U251 (in the Cy5 channel) with LN229 (in the Cy3 channel). Detailed information about the experiments is provided in Table 1. We evaluated the amplification protocol using the criteria described earlier.

### **Enhancement of signal intensity for genes expressed at a low copy number**

We expected the amplification protocol to increase the signal intensity of low expressing genes; in other words, the number of genes having measurable signal intensity above background levels from using the amplification protocol should be higher than that using the regular protocol. We used  $S/N > 2.0$  to determine if a gene has measurable signal intensity on an array; i.e., if the difference between signal intensity and background intensity is greater than 2.0 SD of the local background, then the gene gives adequate signal intensity. We assessed all 9 arrays;

**Table 1: Array Information**

Array ID	Cy5 Channel	Cy3 Channel	Description
R1	U251	LN229	Regular protocol, using 300 µg total RNA
R2	U251	LN229	Regular protocol, using 200 µg total RNA
R3	U251	LN229	Regular protocol, using 100 µg total RNA
R4	U251	LN229	Regular protocol, using 100 µg total RNA
R5	U251	LN229	Regular protocol, using 100 µg total RNA
A6	U251	LN229	Amplification protocol, using 1 µg total RNA, 3 cycles of PCR
A7	U251	LN229	Amplification protocol, using 1 µg total RNA, 3 cycles of PCR
A8	U251	LN229	Amplification protocol, using 1 µg total RNA, 1 cycle of PCR
A9	U251	LN229	Amplification protocol, using 1 µg total RNA, 1 cycle of PCR

**Table 2: Number of Genes (of 2304 Genes) with S/N > 2**

Array	Cy5 Channel	Cy3 Channel
R1	1796 (78.0%)	1801 (78.2%)
R2	960 (41.7%)	1784 (77.4%)
R3	387 (16.8%)	1112 (48.3%)
R4	268 (11.6%)	883 (38.3%)
R5	284 (12.3%)	955 (41.4%)
A6	2097 (91.0%)	1988 (86.3%)
A7	1990 (86.4%)	1729 (75.0%)
A8	1814 (78.7%)	1738 (75.4%)
A9	2038 (88.5%)	1815 (78.8%)

Arrays R1 – R5 were produced using the regular protocol. Arrays A6 – A9 were produced using the amplification protocol.

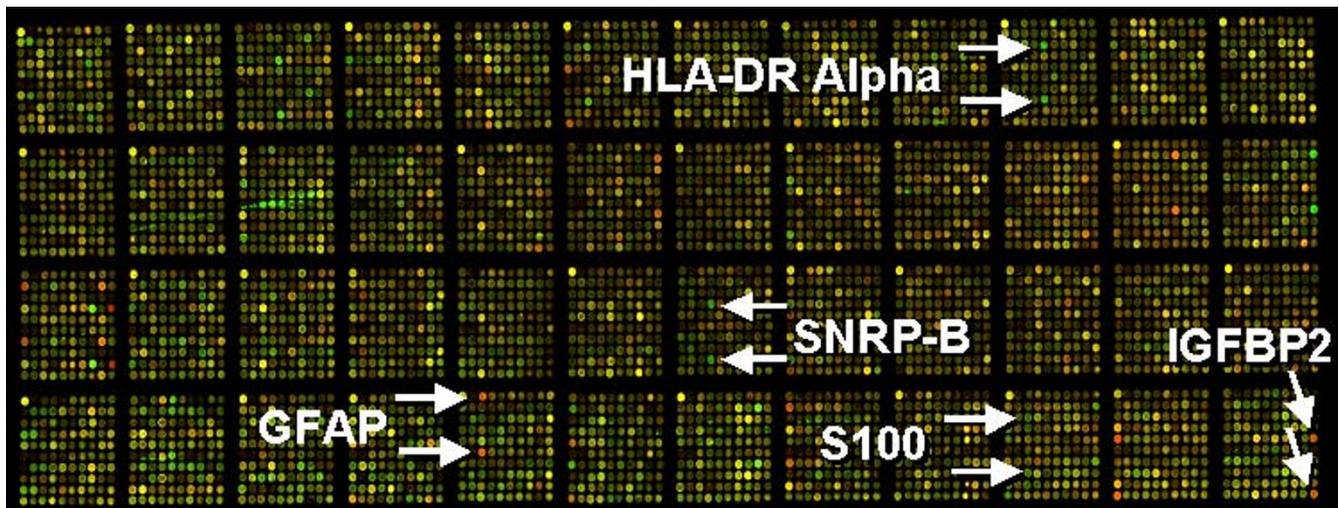
the results are summarized in Table 2. The results showed that the number of genes whose S/N ratio exceeds 2 SD of background is consistently higher in both channels when using the amplification protocol. The signal intensities of low expressors, indeed, are improved so that more genes turned out to be detectable. Our result also showed that the PCR step (1 or 3 cycles) before aRNA amplification had no detectable effect on the final results.

#### **Consistent patterns of differential gene expression**

Finding differentially expressed genes is one of the goals of microarray technology. Ideally, both the regular and amplification protocols would identify the same list of differentially expressed genes. We did not expect to see identical lists of differentially expressed genes generated from the two protocols, since we may not even achieve that goal using the same protocol twice due to the variability associated with microarray experiments [12–14]. However, we do want to see most differentially expressed genes appearing on both lists.

To identify differentially expressed genes, we computed "smooth t-statistics". To apply this method, we first computed the mean log intensity and the standard deviation for the replicated spots within each channel. Since the standard deviation varies systematically with the mean, we then fit a smooth curve representing the standard deviation as a function of the mean. After pooling the smooth estimates of standard deviation from the two channels, we used the pooled estimates to compute a t-statistic for each gene [13,15]. These "smooth t-statistics" can also be viewed as "studentized log ratios"; i.e., as log ratios between channels that have been rescaled to account for the intrinsic variability.

We found that the consistency and reproducibility of smooth t-statistics between experiments is high regardless of the protocol being used. We quantified the reproducibility using the concordance coefficient, which is analogous to the correlation coefficient but measures how well a set of points matches the identity line [16–18]. The con-



**Figure 1**

**Composite array images generated by the amplification protocol.** The labeled cDNA probes were generated from total RNA from the U251 cell line (Cy5) and the LN229 cell line (Cy3), respectively. Seven differentially expressed genes were randomly picked up across the array. Among these, four genes, GFAP, S100, HLA-DR and SNRP-B, were detected by using both the amplification and the regular protocols; three genes, IGFBP2, integrin beta 4 and HLA-A, were detected only using the amplification protocol.

cordance between experiments using the regular protocol ranged from 0.574 to 0.835, with a median of 0.739. The concordance using the amplification protocol ranged from 0.843 to 0.894, with a median of 0.873. The concordance between the experiments using different protocols ranged from 0.562 to 0.796, with a median of 0.652.

Using data from all experiments performed with the regular protocol, we computed a combined set of smooth t-statistics. We computed a similar set of t-statistics using data from all experiments performed with the amplification protocol. Genes are viewed as differentially expressed if the smooth t-statistic exceeds some threshold; for this study, we chose genes with smooth t-statistic greater than four in absolute value. Using this criterion, we found that the regular protocol identified 21 genes that were differentially expressed between U251 and LN229 cell lines, and the amplification protocol identified 28 differentially expressed genes. Fifteen genes were common to both lists. Thirteen genes were found using the amplification protocol but not the regular protocol; the smooth t-statistics for these genes ranged in absolute value from 1.19 to 3.96, with a median of 2.23. Six genes were found using the regular protocol but not the amplification protocol; the smooth t-statistics for these genes using the amplification protocol ranged in absolute value from 3.16 to 3.97 with a median of 3.42. In every case, the sign of the t-scores was the same using the different protocols, indicating that the same cell line overexpressed the gene.

#### **Reproducibility and reliability of signal intensity**

To evaluate whether the amplification protocol preserved signals that were seen using the regular protocol, we first found all the spots in each channel (U251 or LN229) that consistently had  $S/N > 2.0$  in all five experiments using the regular protocol. We then computed the percentage of those genes that also had  $S/N > 2.0$  on each of the arrays using the amplification protocol. We found that 95% to 100% of the genes that are consistently found to be expressed using the regular protocol were also found to be expressed using the amplification protocol. For U251 on four amplified arrays, the percentages are 100%, 99%, 98%, and 100%; for LN229, 100%, 96%, 95%, and 98%.

#### **Confirmation of the results from amplification experiments**

A key criterion for the validity of microarray experiments is whether the results are real and can be confirmed by other approaches. In order to obtain representative information rather than noise from nonspecific binding of targets to the DNA probes in a microarray experiment, the hybridization has to be specific. To demonstrate the hybridization specificity in our microarray experiments, we selectively labeled a specific target such as actin or GAPDH with one dye and hybridized to the microarray. Our results showed that the specific target only hybridized to its corresponding spots on the array [19]. Second, in order to confirm the results, the clones printed on the microarray have to be correct and error free. This is an important issue because most clone libraries used for making microarrays

contain significant errors. To surmount this problem, all the clones printed on our microarrays were sequence verified before printing [20]. Because of those quality control measures, we expected that our microarray results should be confirmed as long as the amplification is stable and consistent among different experiments.

From the results of comparison between U251 and LN229, we selected six genes that showed differential expression either using both the regular and the amplification procedures or only using the amplification procedure. Among these, four genes were detected by both protocols and two genes were only detected by the amplification protocol (Figure 1). Genes that belong to the latter case are most likely expressed at very low levels that fall below the threshold of detection by the regular protocol. These genes were then tested by other molecular methodologies.

We first selected two genes, S100 and GFAP, that are common markers for glial cells. Among U251 and LN229 cells, the microarray results from both protocols showed that S100 was expressed at higher levels in LN229 cells and GFAP at higher levels in U251 cells. U251 and LN229 cells were embedded in paraffin blocks and sectioned. The sections were stained with antibodies for those two antigens. The immunohistochemistry results showed that S100 is expressed at higher levels in LN229 cells and GFAP at higher levels in U251 cells (Fig 2A). Two additional genes, HLA-DR alpha and SNRP-B, that were also obtained using the two protocols, were confirmed by northern blotting (Fig 2A).

We mentioned that some low-expressing genes could only be detected by the amplification protocol and not by the regular protocol. To confirm whether the new information by this approach is reliable, we randomly picked up two differentially expressed genes, IGFBP2 and integrin beta 4, that were identified by the amplification protocol but not by the regular protocol for confirmation using western blotting, northern blotting and immunohistochemistry assays. The results from these experiments showed that all the two genes were expressed at higher levels in U251 cells (Figure 2B) as obtained by the microarray data. It should be noted that for the low expressed genes that could only be detected by the amplification protocol, longer exposure time for the northern blotting is needed and the signal intensity is much lower than that generated from the high expressed genes that could be detected by both protocols (Figure 2B).

In summary, we have carried out a series of confirmation experiments using northern blotting, western blotting and immunohistochemistry assay and all six differentially ex-

pressed genes revealed from the microarray experiments have been confirmed.

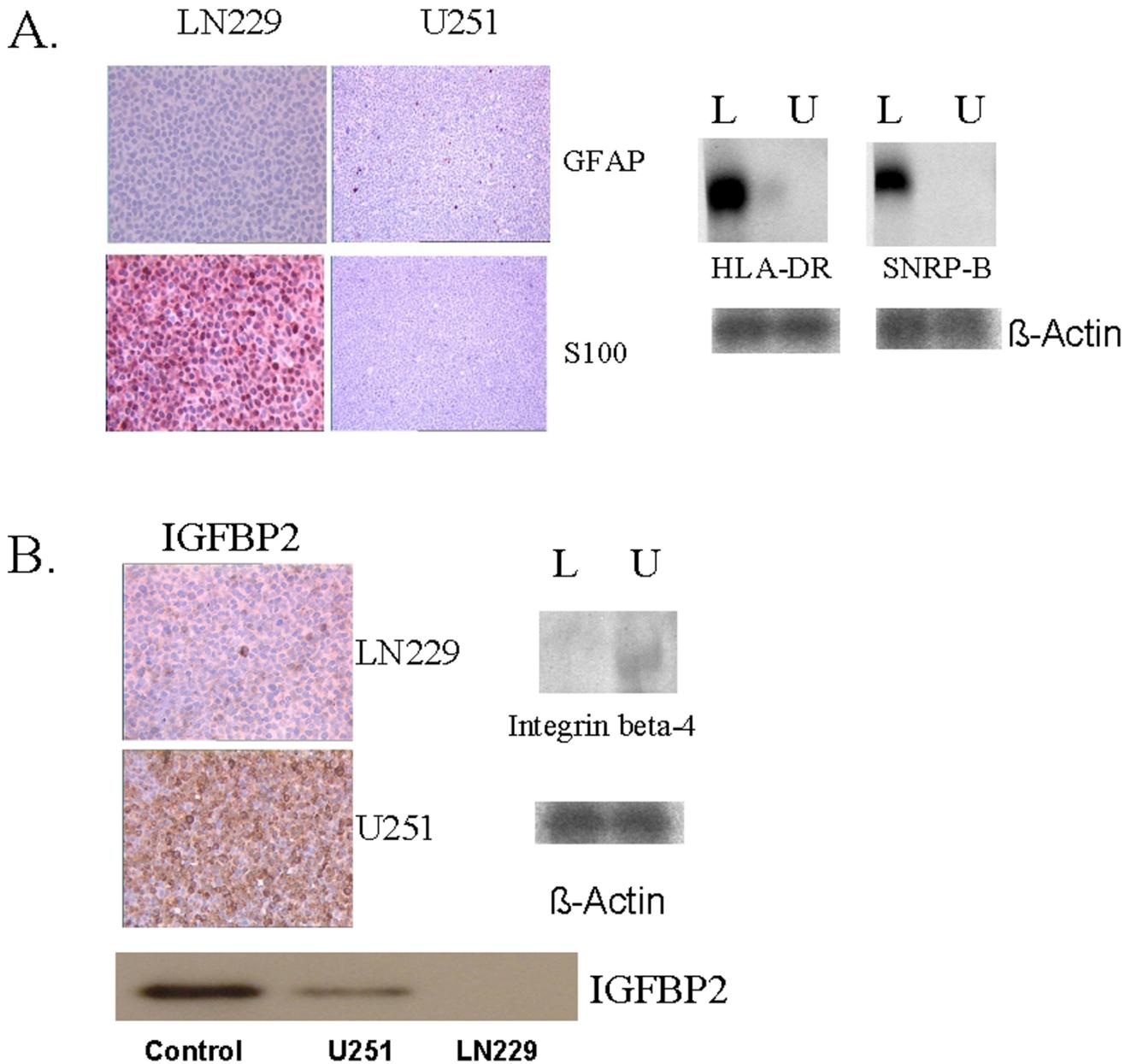
## Discussion

A major difficulty in the study of cancer is that cancer cells are extremely elusive and the genetic mutation events are continuously occurring temporally and spatially. The development of high throughput genomic technologies enables us to screen thousands of genes simultaneously for the informative genes, yet the power of these technologies is greatly attenuated when inappropriate samples are studied. Generally speaking, the most biologically appropriate materials obtained from clinical samples are small in quantity. This observation motivates the development of experimental protocols that allow application of genomic technologies to minute amounts of biological materials. In this manuscript, we have described such a method that allows us to use cDNA microarrays with only 1 microgram of total RNA.

The initial evaluation of the methodology is based on the detection of fluorescent signals on the microarray after hybridization. Although we could see bright spots with the amplification protocols, this apparent "success" does not mean the information so acquired is reliable. False positives are often the result of side-effects during amplification. We set out to assess the accuracy and reliability using both statistical and experimental approaches. Both approaches supported the validity of our protocol in the identification of differentially expressed genes in two cell populations. This demonstrates that the same genes are amplified to the same extent in different cell populations. However, in our results, the amount of amplification varies from gene to gene. Therefore, the amplification process is not balanced and changes the relative transcript levels in a given cell population. Thus, from the amplified results, one cannot infer the relative expression levels of different genes in a single sample. However, in most biological studies, identification of differentially expressed genes among different samples is the main aim. These genes can serve as useful markers for diagnosis, for prediction of therapeutic response, and perhaps as a target for developing new drugs. Thus, we have developed an assay that will enable us to approach a series of important clinical issues given a limited amount of material. This assay has been used in several ongoing studies and useful information has been obtained.

## Conclusion

We have confirmed a powerful and consistent cDNA microarray procedure that can be used to study minute amounts of biological tissue.



**Figure 2**  
**Confirmation of microarray results.** To validate the microarray results seven differentially expressed genes between LN229 and U251 cell lines were selected across the array. **(A)** Four differentially expressed genes detected using both the amplification and the regular protocols were confirmed by immunochemistry assay (GFAP and S100, positive cells are stained brown) and northern blotting (HLA-DR and SNRP-B). **(B)** Three other differentially expressed genes, IGFBP2 and integrin beta 4 detected only using the amplification protocol were confirmed by immunohistochemistry assay (IGFBP2), western blotting (IGFBP2) and northern blotting (integrin beta 4).

**Materials and Methods**

**Microarray production**

A total of 2,304 known human cDNAs were prepared by PCR from the Research Genetics cDNA clone library using the two primers on the vector. The sequences of the two

primers are: up-stream 5'-CTGCAAGGCATTAAGTTGGG-TAAC-3'; down-stream, 5'- GTGAGCGGATAACAATT-TCACACAGGAAACAGC-3'. The PCR products were purified using MultiScreen PCR plates (Millipore Corp., Bedford, MA) and verified by sequencing at our Cancer

Genomics Core Lab prior to printing. The DNA clones, in 394-well plates, were spotted onto poly-L-lysine-coated microscope slides using a robotic arrayer (Genomic Solutions, Ann Arbor, Michigan). All the clones except for the control genes such as GAPDH,  $\beta$ -actin, tubulin and an EST highly similar to GAPDH, are duplicated on the array. After printing, the slides are dried, cross-linked by UV (650 J/cm<sup>2</sup>), washed by water, and stored dry.

#### **RNA amplification and target labeling**

U251 and LN229 human glioma cell lines were maintained in DMEM supplemented with 10% fetal calf serum (Gibco-BRL, Gaithersburg, MD). Total RNA was isolated using TRI reagent (MRC, Cincinnati, OH). 5-Amino-propargyl-2'-deoxycytidine 5'-triphosphate coupled to Cy3 (Cy5) fluorescent dye – Cy 3-AP3-dCTP or Cy5 -AP3-dCTP (Cy3 and Cy5) were purchased from Amersham Pharmacia Biotech Inc. (Piscataway, NJ). Reverse transcription was carried out in a 20  $\mu$ l volume containing 1  $\mu$ g total RNA, 1  $\mu$ g oligdT25- T7 (5'-AAACGACGGCCAGTGAATTGTAATACGACTCACTAT-AGGG -CGATT-3') primer, 1  $\mu$ g template switch primer (5'-AAGCAGTGGTAACAACGC -AGGGACCGGG-3'), 4  $\mu$ l first-strand reaction buffer, 2  $\mu$ l 10 mM dithiothreitol (DTT; Gibco-BRL) 1  $\mu$ l 10 mM dNTPs, 1  $\mu$ l SUPERase-in, (Amstion, Austin, TX) and 200 u Superscript II reverse transcriptase (Gibco- BRL). cDNA synthesis was completed at 42°C for 2 hours. Then 1 u RNase H (Roche, Branchburg, NJ) was added to the reaction followed by incubation at 37°C for 15 min. Full-length ds-cDNA was synthesized by adding 57  $\mu$ l nuclease-free water, 10  $\mu$ l 10 $\times$  PCR buffer (Roche), 10  $\mu$ l 25 mM MgCl<sub>2</sub>, 1  $\mu$ l 10 mM dNTP and 5 u Ampli-Taq Gold DNA Polymerase (Roche). The reaction was carried out at 95°C for 10 min and then 1 or 3 cycles at: 95°C for 1 min., 65°C for 6 min. A prolonged elongation time up to 12 min was used in the final cycle. The PCR product was purified by QIAquick PCR Purification Kit (Qiagen, Valencia, CA). The anti-sense RNA (aRNA) amplification by T7 in vitro transcription was carried out using MEGAscript T7 Kit (Ambion Inc., Austin, TX) following the manufacturer's instructions. After RNA amplification, the DNA template was removed by incubation of the reaction with 4 units of RNase-free DNase I (Ambion) at 37°C for 15 min. aRNA purification was achieved by Rneasy Mini Kit (Qiagen). The purified 5  $\mu$ g aRNA was labeled with Cy3 or Cy5 by reverse transcription in a solution containing 10  $\mu$ g of random hexamer, 4  $\mu$ l first-strand reaction buffer, 2  $\mu$ l 10 mM dithiothreitol (DTT; Gibco-BRL), 1  $\mu$ l of 2 mM dATP, dGTP, dTTP and 1 mM dCTP, 1  $\mu$ l SUPERase-in, (Amstion) and 200 u Superscript II reverse transcriptase (Gibco-BRL). The labeling was completed at 42°C for 2 hours. Then the labeled cDNA was purified using MicroSpin G-50 columns (Amersham Pharmacia). The volume was reduced to about 10  $\mu$ l using

a Speed-Vac system (ThermoSavant, AES2010) before hybridization.

#### **Microarray hybridization and scanning**

To hybridize the arrays, purified and labeled cDNA targets were dissolved in 100  $\mu$ l total volume of ExpressHyb solution (Clontech, Palo Alto, CA) containing 8  $\mu$ g of polydA<sub>40-60</sub> (Amersham Pharmacia), 2  $\mu$ g of yeast tRNA (Gibco-BRL), 10  $\mu$ g of human Cot I DNA (Gibco-BRL). The mixture was heated to 95°C for 10 min, then applied to the slides and covered by a coverslip. Hybridization was carried out at 60°C for 14–16 hours in a moisturized box in an incubator. Slides were washed at 37°C in 1 $\times$  SSC (3 M sodium chloride, 0.3 M sodium citrate), 0.01% SDS, 0.2 $\times$  SSC, 0.01% SDS, and twice in 0.1 $\times$  SSC sequentially for 2 min each. Hybridized arrays were scanned at 10- $\mu$ m resolution on a LSIV scanner (Genomic Solution, Ann Arbor, MI).

#### **Data quantification**

Scanned microarray images (16-bit TIFF formatted files) were quantified with ArrayVision™ (Imaging Research, Inc., St. Catherine's, Ontario, Canada), and values were recorded for spot intensity, local background intensity, and signal-to-noise (S/N) ratio. Spot intensity was computed as the integrated optical density or volume in a fixed-size circle; background intensity was computed as the median pixel value of four diamond-shaped regions at the corners of each spot. Background-corrected intensity was computed by subtracting local background from spot intensity. The S/N ratio was computed by dividing the background-corrected intensity by the standard deviation (SD) of the background pixels.

#### **Data normalization**

Data from all microarrays was imported into S-plus™ (Insightful Corp., Seattle, WA), for further analysis. For each microarray, the background-corrected spot intensities were normalized by setting the 75th percentile equal to 1000. This procedure brings the median log ratio between channels close to 1 for expressed spots, and it also permits us to compare individual channels across arrays. After normalization, any spot whose normalized intensity remained below the threshold value of 150 was considered to be undetectable, and its value was replaced by the threshold value. In our arrays, a normalized value of 150 corresponds roughly to a spot whose S/N ratio equals 1, and thus any spot whose background-corrected intensity falls below this threshold cannot be reliably distinguished from background noise. Finally the background-corrected, normalized signal intensities were log-transformed (base two) for further analysis.

### Northern blotting analysis

The sequencing verified cDNAs used to print the microarrays were used as templates and the probes were labeled using Rediprime II Random Prime Labeling System (Amersham). For northern blotting, 20 µg of total RNA was electrophorized on a denaturing agarose gel, transferred to nylon membrane, and hybridized to <sup>33</sup>P labeled cDNA probes as described previously [11].

### Western blotting analysis

For western blotting, 40 µg of total cellular protein was run on a SDS-PAGE gel, transferred to a nylon membrane, and incubated with IGFBP2 antibody (Santa Cruz Biotechnology, Inc.) following the procedure as described [11].

### Immunohistochemistry assay

Cell pellets were first embedded to paraffin block and sections cut and mounted onto microscope slides. The presence of S100, GFAP, and IGFBP2 antigens was detected by their antibodies following standard immunohistochemistry procedures [11].

### Acknowledgments

This work was partially supported by the Tobacco Settlement Funds as appropriated by the Texas State Legislature, by a generous donation from the Michael and Betty Kadoorie Foundation, by a grant from the Texas Higher Education Coordination Board under the grant number 003657-0039-1999, and by a grant from the RGK.

### References

- International Human Genome Sequencing Consortium: **Initial sequencing and analysis of the human genome.** *Nature* 2001, **409**:860-921
- Hogenesh JB, Ching KA, Batalov S, Su AI, Walker JR, Zhou Y, Kay SA, Schultz PG, Cooke MP: **A comparison of the Celera and Ensembl predicted gene sets reveals little overlap in novel genes.** *Cell* 2001, **106**:413-415
- Chaib H, Cockrell EK, Rubin MA, Macoska JA: **Profiling and verification of gene expression patterns in normal and malignant human prostate tissues by cDNA microarray analysis.** *Neoplasia* 2001, **3**:43-52
- DeRisi J, Penland L, Brown PO, Bittner ML, Meltzer PS, Ray M, Chen Y, Su YA, Trent JM: **Use of a cDNA microarray to analyze gene expression patterns in human cancer.** *Nature Genetics* 1996, **14**:457-60
- Zhang L, Zhou W, Velculescu VE, Kern SE, Hruban RH, Hamilton SR, Vogelstein B, Kinzler KW: **Gene expression profiles in normal and cancer cells.** *Science* 1997, **276**:1268-1272
- Hibi K, Kenji H, Liu Q, Beaudry G, Madden S, Westra W, Wehage S, Yang S, Heitmiller R, Bertelsen A, Sidransky D, Jen J: **Serial analysis of gene expression in non-small cell lung cancer.** *Cancer Research* 1998, **58**(24):5690-5694
- St Croix B, Rago C, Velculescu V, Traverso G, Romans KE, Montgomery E, Lal A, Riggins GJ, Lengauer C, Vogelstein B, Kinzler KW: **Genes expressed in human tumor endothelium.** *Science* 2000, **289**(5482):1197-202
- Phillips J, Eberwine JH: **Antisense RNA amplification: a linear amplification method for analyzing the mRNA population from single living cells.** *Methods* 1996, **10**:283-8
- Wang E, Miller LD, Ohnmacht GA, Liu ET, Marincola FM: **High-fidelity mRNA amplification for gene profiling.** *Nat. Biotechnol* 2000, **18**:457-459
- Lou L, Salunga RC, Guo H, Bittner A, Joy KC, Galindo JE, Xiao H, Rogers KE, Wan JS, Jackson MR, Erlander MG: **Gene expression profiles of laser-captured adjacent neuronal subtypes.** *Nat. Med* 1999, **5**:117-122
- Fuller GN, Rhee CH, Hess KR, Caskey LS, Wang R, Bruner JM, Yung WK, Zhang W: **Reactivation of insulin-like growth factor binding protein 2 expression in glioblastoma multiforme: a revelation by parallel gene expression profiling.** *Cancer Res.* 1999, **59**:4228-32
- Hess KR, Zhang W, Baggerly KA, Stivers DN, Coombes KR: **Microarrays: handling the deluge of data and extracting reliable information.** *Trends Biotechnol* 2001, **19**:463-8
- Baggerly KA, Coombes KR, Hess KR, Stivers DN, Abruzzo LV, Zhang W: **Identifying differentially expressed genes in cDNA microarray experiments.** *J Comp. Biol.* 2001, **8**(6):639-659
- Lockhart DJ, Dong H, Byrne MC, Follettie MT, Gallo MV, Chee MS, Mittmann M, Wang C, Kobayashi M, Horton H, Brown EL: **Expression monitoring of hybridization to high-density oligonucleotide arrays.** *Nat. Biotechnol* 1996, **14**:1675-1680
- Baldi P, Long AD: **A bayesian framework for the analysis of microarray expression data: regularized t -test and statistical inferences of gene changes.** *Bioinformatics.* 2001, **7**:509-19
- Lin L: **A concordance correlation coefficient to evaluate reproducibility.** *Biometrics* 1989, **45**:255-268
- Lin L: **Assay validation using the concordance correlation coefficient.** *Biometrics* 1992, **48**:599-604
- Liao JJZ, Lewis JW: **A note on the concordance correlation coefficient.** *PDA Journal of Pharmaceutical Science and Technology,* **54**:23-26
- Hu L, Cogdell DE, Jia YJ, Hamilton SR, Zhang W: **Monitoring of cDNA microarray production with a common primer target and hybridization specificity with selected targets** *BioTechniques* 2002, **32**:528-534
- Taylor E, Cogdell DE, Coombes K, Hu L, Ramdas L, Tabor A, Hamilton SR, Zhang W: **Sequence verification as quality control step for production of cDNA microarrays.** *BioTechniques* 2001, **31**:5-8

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMedcentral will be the most significant development for disseminating the results of biomedical research in our lifetime."

Paul Nurse, Director-General, Imperial Cancer Research Fund

Publish with **BMC** and your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours - you keep the copyright

Submit your manuscript here:

<http://www.biomedcentral.com/manuscript/>



**BioMedcentral.com**

[editorial@biomedcentral.com](mailto:editorial@biomedcentral.com)