

RNA expression microarrays (REMs), a high-throughput method to measure differences in gene expression in diverse biological samples

Charles E. Rogler*, Tatyana Tchaikovskaya, Raquel Norel, Aldo Massimi¹, Christopher Plescia², Eugeny Rubashevsky, Paul Siebert³ and Leslie E. Rogler

Department of Medicine and Marion Bessin Liver Research Center, ¹Department of Molecular Genetics, Albert Einstein College of Medicine, Bronx, NY, USA, ²Department of Neurosciences, Mt Sinai College of Medicine, New York, NY, USA and ³BD Biosciences-Clontech, Palo Alto, CA, USA

Received April 12, 2004; Revised June 4, 2004; Accepted July 30, 2004

ABSTRACT

We have developed RNA expression microarrays (REMs), in which each spot on a glass support is composed of a population of cDNAs synthesized from a cell or tissue sample. We used simultaneous hybridization with test and reference (housekeeping) genes to calculate an expression ratio based on normalization with the endogenous reference gene. A test REM containing artificial mixtures of liver cDNA and dilutions of the bacterial *LysA* gene cDNA demonstrated the feasibility of detecting transcripts at a sensitivity of four copies of *LysA* mRNA per liver cell equivalent. Furthermore, *LysA* cDNA detection varied linearly across a standard curve that matched the sensitivity of quantitative real-time PCR. In REMs with real samples, we detected organ-specific expression of albumin, *Hnf-4* and *Igfbp-1*, in a set of mouse organ cDNA populations and *c-Myc* expression in tumor samples in paired tumor/normal tissue cDNA samples. REMs extend the use of classic microarrays in that a single REM can contain cDNAs from hundreds to thousands of cell or tissue samples each representing a specific physiological or pathophysiological state. REMs will extend the analysis of valuable samples by providing a common broad based platform for their analysis and will promote research aimed at defining gene functions, by broadening our understanding of their expression patterns in health and disease.

INTRODUCTION

Evolutionary selection pressure functions both at the organismal level, and at the molecular level, by precisely tuning the regulatory properties of enhancers and promoters, so that each gene product is produced when and where it is needed and in sufficient quantities to supply its required function (1–3). Consequently, the temporal and spatial pattern of expression of a gene is a catalog of biological processes in which a gene can

have a vital function. Using cDNA microarrays, researchers can simultaneously measure steady-state mRNA levels in all the known genes and thousands of expressed sequence tags (ESTs) expressed in a cell (4–6).

Genome-wide expression analysis has advanced transcriptional based research in all areas of biology. In mammalian biology for example, cDNA microarray approaches have identified novel genes involved in the cell cycle (7,8), specific differentiation programs (9,10) and specific disease states (11,12) to mention just a few. In cancer biology, an important application has been the identification of distinct subtypes of tumors, such as subtypes of breast tumors (13,14), lymphomas (15), kidney tumors (16), melanomas (17) and other tumor types (18).

In all the above cases, the genome-wide scans have led to the identification of candidate disease-specific genes in defined sets of samples. Follow-up studies, on the expression of the best candidate genes in a much larger sample base, are then needed to quantitate and validate their involvement in specific biological processes or diseases. Experimental approaches to investigating the broader ‘expression niche’ of candidate genes include, among others, the use of tissue northern blots, RNA dot blots (19) and tissue microarrays (20). Acquisition of samples and sample processing for large sets of samples are often the rate-limiting step in this process. Therefore, there is a need for an experimental tool that provides large sets of samples that can be probed for the expression of specific genes in a high-throughput manner. In addition, the tool should use small amounts of valuable samples so that the effective use of those samples can be extended.

In this report we describe the development and use of a new microarray technology, called RNA Expression Microarrays (REMs), that addresses the above needs. REMs are produced by spotting cDNAs synthesized from the poly(A)⁺ mRNAs of a tissue. REMs have the advantage of precise internal normalization, quantitative comparisons between the samples and a capacity for high-throughput analysis of thousands of diverse samples simultaneously. We have validated the technology using artificial mixtures and compared it with the leading quantitative expression analysis method, and applied REM technology toward biologically relevant questions in developmental and cancer biology.

*To whom correspondence should be addressed. Tel: +1 718 430 2607; Fax: +1 718 430 8975; Email: crogler@acom.yu.edu

METHODS

Preparation of fluorescent probes

Gene-specific sense and antisense primers ~500 bp apart are identified near the 3' end of the cDNA sequence of selected gene. A T7 promoter sequence is attached to the antisense primer and the cDNA fragment is PCR amplified, purified using the Qiaquick PCR purification kit and the product is sequence verified. An antisense RNA is synthesized using T7 RNA polymerase according to the Epicentre AmpliScribe T7 Flash transcription kit protocol, (Epicentre Cat. no. ASF3257), except that the reaction is carried out at 42°C for 1 h. The antisense RNA is purified using a RNeasy Mini kit from Qiagen (Cat. no. 74104). Five micrograms of antisense RNA, at 0.3 µg/µl in H₂O, is annealed at 70°C for 5 min with 6 µM sense primer. After annealing, Cy3- or Cy5-labeled sense strand cDNA is synthesized using 10 U/µl Invitrogen Superscript III reverse transcriptase, in 50mM Tris-HCl, pH 8.3, 75 mM KCl, 3 mM MgCl₂, 10 mM DTT, 1 U/µl RNaseOUT (Invitrogen, Cat. no. 10777-019), 500 nM dATP, dCTP, dGTP and 200 nM dTTP, plus either Cy3- or Cy5-labeled dUTP at 100 nM, at 50°C for 2 h (reaction volume normally 40 µl). After completion of the reaction, an equal volume of 17.5 mM MgCl₂ and 250 mM Tris-HCl, pH 7.4, containing 4 U of RNase H, is added and incubated for 30 min at 37°C, followed by the treatment with 0.5 U/µl of RNase 1 and RNase 1 buffer (10 mM Tris-HCl, pH 7.5, 5 mM EDTA and 200 mM sodium acetate (Promega no. 4261) for 10 min at 37°C. Probe solutions containing either Cy3 or Cy5 are combined and purified together using a Qiaquick PCR purification kit. Final purification is accomplished by elution from the Qiagen columns using 10 mM Tris-HCl (pH 8.5) as elution buffer (Qiagen protocol), the combined probes are precipitated by adding one-third volume of 7.5 M ammonium acetate, followed by 2.5 vol of absolute ethanol, and precipitation at -80°C for 20 min. The precipitates are collected by centrifugation at 13 K in a microcentrifuge for 15 min and the pellets are washed with 75% ethanol and air-dried.

Immediately before REM hybridization, the pellet containing the combined Cy3- and Cy5-labeled probes is dissolved in 20 µl of hybridization buffer containing 35% formamide, 0.5% SDS, 2.5× Denhardt's solution, 4× SSPE, 0.2 µg/µl yeast tRNA, 0.1 µg/µl poly(dA) and 2.5 µg/µl mouse/human Cot 1 DNA. The probe is boiled at 95°C for 2 min, snap-cooled, spun down in a microcentrifuge at 13 K for 5 min and pre-hybridized at 50°C for 1 h.

Preparation of REMs for hybridization

Dust from the slide is removed with air from a Fisherbrand super friendly Air'IT (Cat. no. 23-022523). The array face of the REM is moisturized over boiling water for 5 s and the DNA is immediately crosslinked to the slide with 250 mJ of UV irradiation in a Bio-Rad UV GS GENE LINKER. The slide is re-moisturized over steam for 5 s and placed (array side up) on a 100°C hot plate for 3–5 s. Then the slide is rinsed in 0.1% SDS for 10–20 s, followed by ddH₂O for 10–20 s and then incubated at 95°C in ddH₂O for 3–5 min. The slide is dipped in absolute ethanol and excess ethanol is removed by centrifugation in a 50 ml tube at 1000 r.p.m. for 4 min. The slide is placed, array side up, in a microarray slide hybridization

chamber. Approximately 20 µl of prehybridization solution (prehybridization solution is 35% formamide, 4× SSPE, 0.5% SDS, 2.5× Denhardt's and 0.2 µg/ml salmon sperm DNA) is added over the arrayed samples and the coverslip is placed over the samples avoiding bubbles. The slide is incubated in the hybridization chamber that is humidified by adding 10 µl of water in each corner, for 1–2 h at 50°C. After incubation the coverslip is removed by dipping in water, the slide is dried by centrifugation as above, dust is removed as described, the slide is returned to the chamber, covered by hybridization solution containing a mixture of Cy3- and Cy5-labeled probes and by coverslip, and incubated in humidified hybridization chamber for 16–20 h at 50°C.

After hybridization, the cover slip is removed by immersing REM in 100 ml of 2× SSC/0.1% SDS then washed with several hundred milliliters of 0.2× SSC/0.1% SDS with stirring for 10–15 min at room temperature, washed with 0.2× SSC and then with 0.1× SSC for every 15 min. Slide is dried by centrifugation as above, stored at room temperature, in dark, until scanning (preferably the same day).

Preparation of single-stranded LysA antisense cDNA and dilution into antisense liver cDNA

A 1 kb segment of a bacterial clone for diaminopimelate decarboxylase (LysA, ATCC accession number 87482) was subcloned into pBluescript II KS+. The clone contained a 60 nt artificial poly(A) tail at its 3' end. A 1.1 kb DNA fragment was amplified from the plasmid using antisense T7 and sense T3 primers homologous to plasmid sequences and the PCR product was sequence verified. Sense strand LysA aRNA was synthesized using an Ambion MEGAscript T3 RNA polymerase kit (Cat. no. 1338). Antisense cDNA was synthesized from the full-length aRNA using an oligo(dT) primer and Superscript II reverse transcriptase followed by removal of the RNA template with RNase 1 and the purification of single-strand antisense cDNA over a Qiagen PCR purification column (Cat. no. 28104). Purified products were measured by OD₂₆₀, checked for correct size.

A large batch of single-stranded liver cDNA was synthesized from 2.5 mg of total RNA from a C57/B16 female mouse and used as the carrier for all the LysA dilutions. LysA 1.1 kb antisense LysA cDNA was mixed with liver cDNAs at 12 levels each representing a 2-fold dilution of LysA per liver cell cDNA equivalent. Our mixtures were based on a 50 µg/ml solution of 1000 bp segment of single-stranded DNA containing 9.1×10^{13} molecules of DNA per milliliter (21). Based on the above standard, we made series dilutions in which LysA was varied from 9000 to 4 copies per liver cell equivalent. All mixtures were prepared in 3× SSC solution. Mixtures were also based on 0.2 pg mRNA per liver cell.

Preparation of LysA sense Cy dye-labeled probe for REM hybridization

A set of nested primers were used to generate a 533 bp sub-fragment of LysA from the 1.1 kb antisense cDNA produced above. The gene-specific primers for this PCR fragment were as follows: 5'-CGAGCAAAGCATTCTCATCA-sense and 5'-T7 linked antisense primer TAATACGACTCACTATAG-GGCTCCTCCAAGATTCAGCAC. T7 RNA polymerase was used to generate an antisense LysA aRNA, the 533 bp

fragment [this aRNA does not contain either oligo(dT) or T7 polymerase promoter sequences]. The final, ~513 bp, sense strand Cy dye-labeled LysA probe was synthesized from 5 µg antisense aRNA using the sense strand primer, at 30 pmol, and reverse transcriptase in the standard probe synthesis conditions described above.

Quantification of LysA by real-time PCR

TaqMan probe and primers were designed with Primer Express Software (Applied Biosystems) and synthesized by Operon (Qiagen) as follows: 5'-GAAACGGGTCCTC-CATCGA-forward primer; 5'-AGTCATGCGTATGCGCT-TCTAC-reverse primer; and 5'-6FAM-TTCTTCTTCGGA-TCACGCCCGG-TAMRA-probe. The TaqMan Rodent Gapdh Control Reagents containing VIC-labeled probe and primers (P/N 4308313; Applied Biosystems) were used to quantify a reference gene expression. Series dilutions of LysA cDNA mixed with mouse liver cDNA were prepared in a way such that a particular reaction mostly contained the same amount of corresponding dilution that was printed on the REM slide. TaqMan Universal Master Mix (P/N 4304437) was used to prepare reaction mixtures containing 900 nM of each primer and 250 nM of appropriate TaqMan probe. We performed a single gene reaction for LysA or Gapdh in each well. For each data point, we had three repetitions and used 96-well optical PCR plates (P/N 4306737; Applied Biosystems). The plates were sealed, spun down and reactions run in an ABI PRISM 7000 Sequence Detection System under default conditions: 50°C for 2 min, 95°C for 10 min, and 40 cycles of 95°C for 15 s and 60°C for 1 min.

Primers for specific gene probes synthesis

All primers were selected using Primer 3 public available program which can be found at http://www.broad.mit.edu/cgi-bin/primer/primer3.cgi/primer3_www.cgi. All suggested primers sequences were double checked for gene specificity using available gene databases. The following genes and primers were selected.

MYC gene probe (NM_002467 *Homo sapiens* v-myc myelocytomatosis viral oncogene homolog) (avian), 5'-AGAG-AAGCTGGCCTCCTACC-forward, 5'-T7 (GTAATACGACTCACTATAGGG)GCCTCTTGACATTCTCCTCG-reverse, product size 632 bp.

GP gene probe (X58295 plasma glutathione peroxidase 3), 5'-CATCTGACCGCCTCTTCTGG-forward, 5'-T7 (GTAAT-ACGACTCACTATAGGG)CATCTGACCGCCTCTTCTGG-reverse, product size 308 bp.

ACTB gene probe (X00351 *Homo sapiens* cytoplasmic beta-actin), 5'-CTACGTCGCCCTGGACTTTCGAGC-forward, 5'-T7, (GTAATACGACTCACTATAGGG)GATGGAGCC-GCCGATCCACACGG-reverse, product size 384 bp.

B2M gene probe (NM_004048 *Homo sapiens* beta-2-microglobulin), 5'-GTGCTCGCGTACTCTCTCT-forward, 5'-T7 (GTAATACGACTCACTATAGGG)ACCTCTAAG-TTGCCAGCCCT-reverse, product size 578 bp.

A 23 kDa highly basic protein (X56932 *Homo sapiens* ribosomal protein L13A) (RPL13A), 5'-TAAACAGGTACTG-CTGGGCCGGAAGGTG-forward, 5'-T7 (GTAATACGACTCACTATAGGG)CACGTTCTTCTCGGCCTGTTTCCG-TAGC-reverse, product size 483 bp.

Alb1 gene probe (NM_009654 mouse albumin 1), 5'-GA-CAAGGAAAGCTGCCTGAC-forward, 5'-T7 (GTAATAC-GACTCACTATAGGG)AGTTGGGGTTGACACCTGAG-reverse, product size 750 bp.

Gapdh gene probe (NM_008084 mouse glyceraldehyde-3-phosphate dehydrogenase), 5'-AACTTTGGCATTGTGGAA-GG-forward, 5'-T7 (GTAATACGACTCACTATAGGG)-TGTGAGGGAGATGCTCAGTG-reverse, product size 599 bp.

Primer sequences for UBI, Hnf4 and Igfbp1 are available upon request to CER. All specific PCR products were sequence-verified and used as templates for antisense RNA synthesis by *in vitro* transcription followed by labeling procedure as described above.

Quantitative real-time PCR assay for human MYC

Template. Samples of individual tumor or normal SMARTTMcDNAs (22–24), were diluted to obtain template amounts per reaction of 100, 200, 400 or 800 pg. These amounts matched the amount of SMART cDNA printed on the REM. For the five pairs of lung tumor/normal SMART cDNAs, we tested three replications of each sample at the 400 pg per spot level and for the amplification efficiency two individual tumor/normal samples in all four different concentrations were tested too.

TaqMan primers and probes. Assay on demand gene expression reagents were from Applied Biosystems. Each assay consisted of forward and reverse primers and MGB (Minor Groove Binder) probe with 6FAM at the 5' end and non-fluorescent quencher at the 3' end mixed in 20× dilutions. TaqMan Universal PCR master mix (P/N 4304437; Applied Biosystems) was diluted 2-fold with water and appropriate amount of assay mixture, and aliquots of 20 µl were dispensed into wells on the reaction plate (P/N 4306737; Applied Biosystems). An aliquot of 5 µl, containing designated amounts of SMART cDNA were added to the reaction mixtures. Target gene and reference gene assays were run as single reactions on the same plate. The following assays were used:

Hs00153408_m1 for MYC oncogene (NM_002467); 5'-GCAGCGACTCTGAGGAGGAACAAGA, reporter position is between exon 2 and 3;

Hs00187842_m1 for beta-2-microglobulin (NM_004048), forward primer 5'-AGGCTATCCAGCGTACTCCAAAGAT, reporter position is between exon 1 and 2;

Hs99999903_m1 for beta-actin (GenBank mRNA X00351), forward primer 5'-TCGCTTTGCCGATCCGCCGCCCGT, reporter position is at exon 1.

Synthesis of cDNAs for printing

Total RNA was isolated using the Qiagen RNA purification procedure (Qiagen no. 75144) according to the manufacturer's instructions. RNA quality was monitored using an Agilent 2100 bioanalyzer (LabChip, Caliper Technologies Corp.). Invitrogen Superscript III reverse transcriptase (Cat. no. 180080-044) was used to synthesize cDNA from 100 µg of total RNA using an Oligo dT primer. After synthesis was completed, the samples were heated at 94°C for 2 min, and then treated with 0.5 U/µl of RNase I and RNase I buffer [10 mM Tris-HCl (pH 7.5), 5 mM EDTA and 200 mM sodium acetate] (Promega no. 4261) for 10 min at 37°C.

Single-stranded cDNA was separated using the Qiaquick PCR purification protocol (Qiaquick Spin Handbook, p. 18), except that an additional 35% guanidine hydrochloride wash step was included after binding cDNA to the Qiaquick column. cDNA was eluted with 10 mM Tris-HCl, pH 8.5, and precipitated with one-third volume of 7.5 M ammonium acetate and 2.5 vol of absolute ethanol. cDNA was pelleted, washed with 75% ethanol and dissolved in water. Concentrations were adjusted to 100, 200, 400, or 800 ng/ μ l in 3 \times SSC for printing. cDNA quality was monitored by running samples on 1% agarose gels and checking for a smear of cDNAs from ~500–3000 bases in length.

AECOM microarray printing procedure for REM microarrays

The REM microarrays were produced with the custom built microarray printer at the AECOM Microarray Facility. Details of the equipment can be viewed on our website (<http://microarray1k.aecom.yu.edu/>). Following is the printer configuration and parameters used for printing.

Printhead and pins. Telechem SPH48 printhead with pins spaced 4.5 mm center-to-center, populated with 16 split-tip

pins, part no. SMP3, arranged in a 4 \times 4 array, each producing a nominal 100 μ m diameter spot.

Dot spacing. Each of the 16 pins forms a domain which was programmed to generate a uniformly spaced 12 \times 12 square dot pattern, with a center-to-center dot spacing of 365 μ m.

Printing parameters. The printing program was configured to produce four replicates of each sample for every microscope slide. This subdivides each domain area into four subdomains containing 3 \times 12 unique dots. With each pickup, each pin produces four equally spaced spots per domain, one each per subdomain, from the same sample. The on-slide dwell time was 100 ms while the HEPA filtered environment was maintained at 25°C and 50% RH.

Microscope slides. The substrate used was the Corning GAPS II amino silane coated slides.

RESULTS

REMs are a reverse format microarray, in which the high-complexity 'target' is bound to a solid support and labeled probes from at least two genes are hybridized simultaneously to the microarray (Figure 1). The cDNA printed on glass

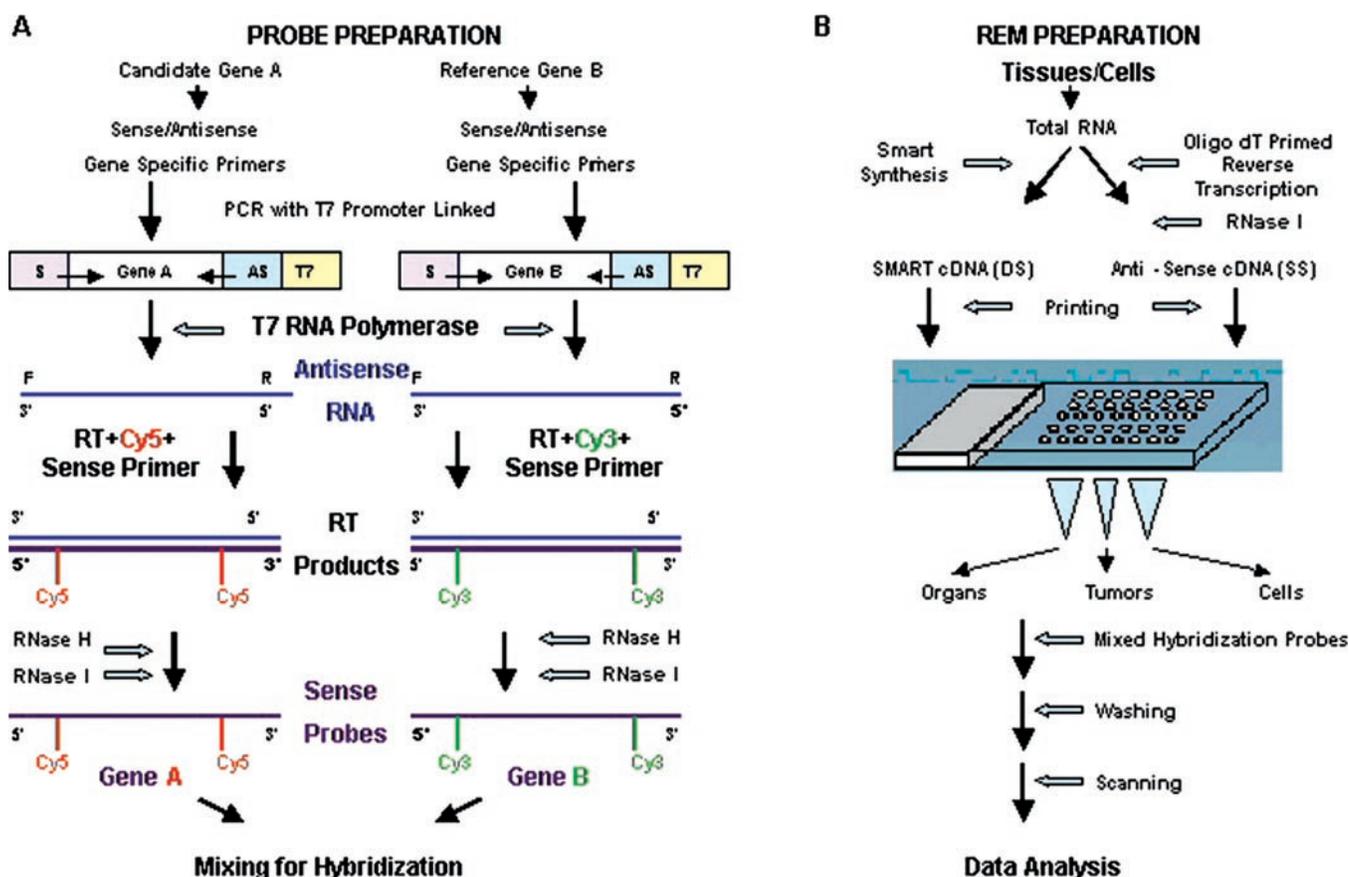


Figure 1. Overview of REM technology. (A) Illustrates the probe preparation protocol starting with PCR using forward and reverse gene-specific primers linked to a T7 promoter. T7 RNA polymerase produces an antisense RNA (blue line), and then reverse transcriptase produces a sense strand Cy3- or Cy5-labeled cDNA probe (magenta plus red or green Cy dye). The Cy3 and Cy5 probes are made single-stranded with RNases and then mixed prior to REM hybridization. (B) REM production and processing. Total RNA from tissues or cells serves as template for reverse transcriptase to synthesize a cDNA primed by Oligo(dT) or a SMART cDNA. cDNAs are printed on Corning GAP slides. High stringency hybridization is carried out with mixed Cy3 and Cy5 probes, in a humidified hybridization chamber followed by washing, scanning and processing of the data using custom made scripts in a Linux operating system.

microscope slides can either be single-stranded antisense cDNA produced by reverse transcription, or the cDNA can be rendered double-stranded by amplification using SMARTTM DNA technology (22–24). The kinetics of hybridization are assumed to be similar to that of RNA dot or northern blots in which the cloned probe is in hybridization excess around the complex mRNA (cDNA) that is bound to the solid support.

During the development of the technology, we tested several different solid supports and different densities of sample printing per spot. We tested total RNAs, poly(A)⁺ RNAs, A RNAs (25), single-stranded cDNAs and SMART cDNAs (22–24). While RNA samples were successfully hybridized, they were very sensitive to RNase degradation, and cDNAs were found to be much more stable substrate for printing and hybridization. Therefore, in this report we only present data using our current REM protocol that involves printing cDNAs on silane-coated glass microscope slides at cDNA spotting densities representing ~1000–4000 cell equivalents per spot. REMs are hybridized simultaneously with a test probe, usually labeled with Cy5 (red) fluorescent dye and a housekeeping gene probe, usually labeled with a Cy3 (green) fluorescent dye (Figure 1). Hybridization signals are measured with a laser scanner (26), and fluorescence data are processed using gene pix software (Axon, Garden City, CA). Data sorting and analysis are carried out using customized computer scripts written using a Linux operating system, and plotted using *Gnu Plot* software.

Organ-specific gene expression detected with REMs

Our initial test of REM technology was to determine whether we could detect organ-specific hybridization of test probes. We decided to use albumin as an example of an abundant liver-specific probe, Hnf4 as a liver preferential transcription factor probe and insulin-like growth factor binding protein-1 (Igfbp-1) as a gene weakly expressed in the liver. Cy5-labeled albumin, Hnf4 and Igfbp-1 probes were synthesized along with Cy3-labeled Gapdh probe. REMs were produced by printing single-stranded antisense cDNAs from a set of mouse organs at a density per spot that represented cDNA from 4000 cells (21). In the case of liver, we printed cDNAs from six different livers, representing one CD1 male, two C57Bl/6 males, and one CD1 female and two C57Bl/6 females. Each cDNA sample was printed in quadruplicate. Thus, the overall ratio of albumin/Gapdh for liver was calculated from 48 quantitative fluorescence measurements (six liver samples, 4× spotting and two probes simultaneously hybridized). Other mouse organs were also represented by multiple samples and each was also quadruplicate spotted.

We hybridized a REM-containing cDNAs from 25 mouse organs with Cy5-labeled albumin plus Cy3-labeled Gapdh probes (Table 1). The hybridization revealed strongly red spots for liver and green spots for all the other organs, as expected. Examples of hybridized spots, viewed as the combined Cy5–Cy3 computer image, are shown in Figure 2A. Using a customized computer script in Linux, we calculated the ratio of albumin signal versus Gapdh for the entire set of mouse organs (Table 1). The ratio for albumin was 10.76 ± 4.08 for liver, whereas the average ratio for the other organs was 0.25 ± 0.1 , clearly demonstrating the strong liver-specific expression.

Table 1. Quantitative analysis of albumin, Hnf-4 and Igfbp-1 expression in mouse organs and Gapdh hybridization to murine organ cDNAs printed on a REM at 400 pg/spot

| Tissue | Ratio gene/Gapdh \pm SD | | |
|----------------|------------------------------------|-----------------------------------|-----------------------------------|
| | Albumin | Hnf4 | Igfbp1 |
| Adipose | 0.17 \pm 0.02 | 0.34 \pm 0.03 | 0.12 \pm 0.01 |
| Adrenal gland | 0.36 \pm 0.2 | 0.42 \pm 0.03 | 0.14 \pm 0.02 |
| Bladder | 0.45 \pm 0.45 | 0.46 \pm 0.1 | 0.15 \pm 0.01 |
| Brain | 0.15 \pm 0.03 | 0.36 \pm 0.07 | 0.1 \pm 0.02 |
| Cerebellum | 0.18 \pm 0.03 | 0.36 \pm 0.04 | 0.12 \pm 0.01 |
| Colon | 0.23 \pm 0.06 | 0.43 \pm 0.08 | 0.11 \pm 0.04 |
| Duodenum | 0.2 \pm 0.06 | 0.4 \pm 0.1 | 0.12 \pm 0.02 |
| Epididymis | 0.23 \pm 0.02 | 0.53 \pm 0.01 | 0.16 \pm 0.01 |
| Heart | 0.13 \pm 0.03 | 0.33 \pm 0.05 | 0.15 \pm 0.06 |
| Intestine | 0.17 \pm 0.02 | 0.38 \pm 0.05 | 0.09 \pm 0.02 |
| Kidney | 0.22 \pm 0.07 | 0.42 \pm 0.08 | 0.12 \pm 0.03 |
| Liver | 10.76 \pm 4.08 | 1.04 \pm 0.23 | 0.14 \pm 0.04 |
| Lung | 0.38 \pm 0.11 | 0.47 \pm 0.07 | 0.2 \pm 0.08 |
| Mammary | 0.58 \pm 0.47 | 0.43 \pm 0.01 | 0.1 \pm 0 |
| Muscle | 0.1 \pm 0.1 | 0.26 \pm 0.08 | 0.04 \pm 0.03 |
| Ovary | 0.23 \pm 0 | 0.42 \pm 0.01 | 0.13 \pm 0.03 |
| Pancreas | 0.22 \pm 0.02 | 0.42 \pm 0.06 | 0.51 \pm 0.09 |
| Penial gland | 0.22 \pm 0.02 | 0.43 \pm 0 | 0.17 \pm 0.01 |
| Salivary gland | 0.29 \pm 0.03 | 0.36 \pm 0.09 | 0.22 \pm 0.02 |
| Skin | 0.19 \pm 0.1 | 0.34 \pm 0.12 | 0.11 \pm 0.07 |
| Spleen | 0.24 \pm 0.05 | 0.45 \pm 0.1 | 0.42 \pm 0.19 |
| Stomach | 0.29 \pm 0.1 | 0.5 \pm 0.06 | 0.15 \pm 0.05 |
| Testis | 0.25 \pm 0.04 | 0.5 \pm 0.03 | 0.15 \pm 0.01 |
| Thymus | 0.2 \pm 0.03 | 0.41 \pm 0.01 | 0.11 \pm 0.01 |
| Uterus | 0.21 \pm 0.09 | 0.42 \pm 0.03 | 0.14 \pm 0.04 |

Bold/italics show highest ratios for the test gene.

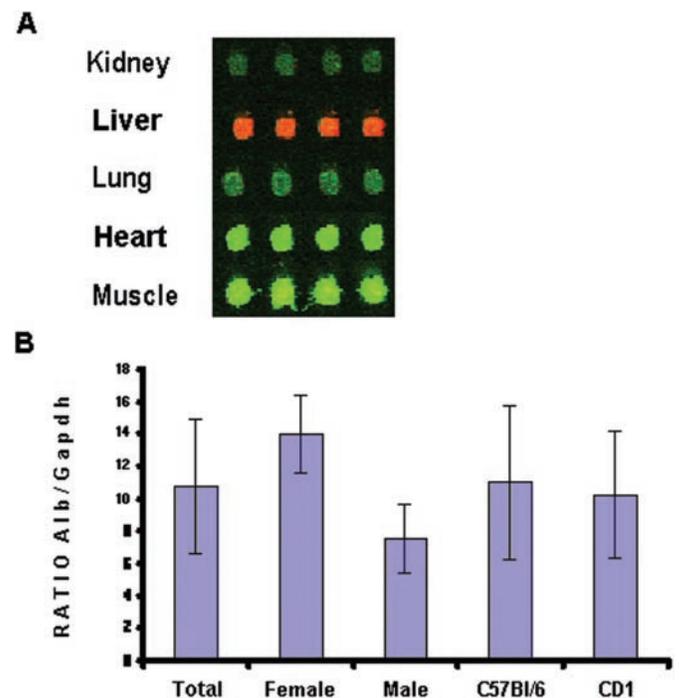


Figure 2. (A) Organ-specific hybridization of albumin to liver cDNA. A mouse organ REM was hybridized with Cy 5-labeled albumin and Cy3-labeled Gapdh probes. The combined computer image shows red for liver cDNA spots and green spots for all other organs. (B) Sorting of liver-specific albumin expression according to sex and genotype of donor mouse.

However, the standard deviation for liver-specific hybridization was very high, suggesting an unexpectedly high level of variability in albumin expression between the different liver samples. To investigate this, we sorted the liver data according to sex, or genotype, of the mice from which the liver samples were taken. By this analysis, we observed no difference due to mouse genotype; however, we observed a significant increase in albumin expression in female liver cDNAs (Figure 2B). This makes sense biologically since an important function of albumin is as a serum carrier protein for estrogen in females (27). These data demonstrate that the REM technology can reveal new information about gene expression differences due to sex and/or genotype.

A second mouse organ REM was hybridized with *Hnf4* plus *Gapdh*. Analysis of the hybridization again revealed liver preferential expression as expected along with expression in all other organs tested (Table 1). Hybridization of a third REM with *Igfbp-1* and *Gapdh* probes revealed the strongest *Igfbp-1* hybridization in the pancreas and spleen, in contrast to liver expression (Table 1).

Normalization with housekeeping genes

It is not possible to be certain that every spot on a REM is equally loaded with cDNA. Therefore, as stated above, it is essential to have an internal housekeeping gene control. The data in Table 2 were generated from a separate REM printing, in which several cDNA samples including muscle and brain were inadvertently overloaded. Thus, for example, albumin hybridization was very high in the muscle cDNA. However, after normalization with *Gapdh*, the ratio for albumin expression versus *Gapdh* in muscle is very low (0.08) as was expected. These data demonstrate that the test gene expression can be accurately normalized against an internal reference for quantitative analysis even when spots contain highly variable amounts of cDNA.

Standard curve generated with artificial mixtures

We next set up an experiment to determine the accuracy of REM technology for detecting rare transcripts in a complex liver cDNA mixture. We prepared artificial cDNA mixtures in which we spiked a liver cDNA preparation with various levels of the bacterial gene, diaminopimelate decarboxylase (*LysA*; ATCC accession number 87482). These mixtures were printed on silane-coated glass microscope slides at 400 or 800 pg of liver cDNA per spot. An aliquot of 400 pg cDNA represents the cDNA from approximately 2000 hepatocytes and the levels of spiked *LysA* cDNA ranged from approximately

9000 copies per cell equivalent (i.e. 1.8×10^7 copies per 400 pg of sample) to approximately two copies per cell equivalent (4×10^4 copies per 400 pg of sample) (21). The results from a set of standard mixtures, printed in quadruplicate and hybridized simultaneously with a green (Cy3) *LysA* probe and a red (Cy5) *Gapdh* reference probe are shown in Figure 3A. The computer combined image shows that spots containing the high level of *LysA* are green and those with a low or undetectable *LysA* level are red, representing solely *Gapdh* reference gene hybridization. A dye reversal experiment revealed a reversed pattern of colored spots, demonstrating the accuracy and reproducibility of the hybridization and detection technology (Figure 3B).

Quantitative analysis of the hybridization signals from Figure 3A allowed us to calculate a ratio for the *LysA* gene versus *Gapdh* across the standard curve of 400 pg spots (Figure 4). These data showed an increasing ratio from 4 to 9000 copies of *LysA* per liver cell cDNA equivalent. A duplicate set of mixtures printed at 800 pg per spot produced a standard curve that was virtually identical to that obtained with the 400 pg per spot series (Figure 4). Therefore, the ratio of test gene versus reference gene is independent of the density of spotting.

Comparison with real-time quantitative PCR

We compared REM technology to quantitative real-time PCR by analyzing six of the standard mixtures from the above analysis by both technologies. The real-time PCR data, expressed as a negative log of the *C_t* value (28), and the REM data, expressed as the log 2 of the *LysA/Gapdh* ratio, are plotted on the y-axis in Figure 5. The two data sets are compared across a set of known amounts of *LysA*, expressed as the log 2 of the *LysA* copy number per reaction or spot. This plot shows a striking parallel from an abundance of

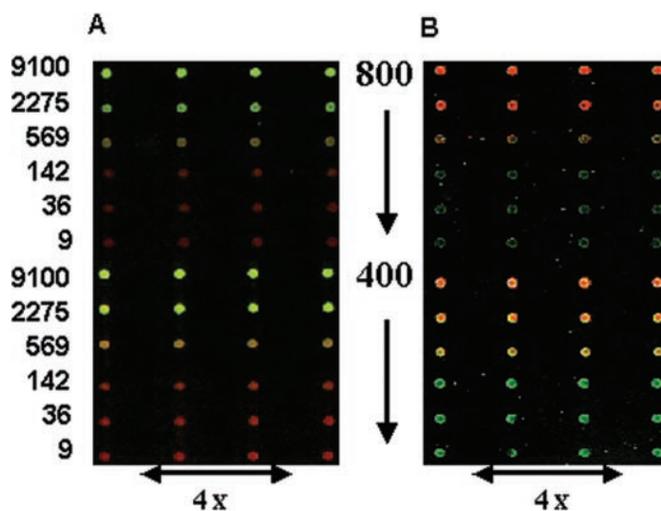


Figure 3. Hybridization of a set of standard liver cDNA mixtures containing increasing amounts of bacterial *LysA* antisense cDNA. (A) *LysA* abundance varying from approximately 9 to 9100 copies *LysA* cDNA per cell equivalent are shown in this figure (left vertical labels). Mixtures of cDNAs were printed at 400 pg/spot (right vertical labels). *LysA* was labeled with Cy3 (green) and *Gapdh* with Cy5 (red), and the green image corresponds to high *LysA*. (B) Dye reversal experiment, *LysA* was labeled with Cy 5 and *GAPDH* with Cy 3, and high *LysA* is a red image.

Table 2. Data demonstrating effectiveness of internal normalization approach for calculating albumin expression in murine organs

| Organ | <i>Gapdh</i> Intensity | Albumin Intensity | ALB <i>Gapdh</i> (ratio) |
|-----------|------------------------|-------------------|--------------------------|
| Kidney | 781 ± 53 | 259 ± 19 | 0.33 ± 02 |
| Liver | 743 ± 53 | 3971 ± 229 | 5.36 ± 0.26 |
| Lung | 1153 ± 82 | 392 ± 28 | 0.34 ± 0.04 |
| Brain | 16 272 ± 1093 | 2808 ± 284 | 0.17 ± 0.01 |
| Intestine | 2322 ± 38 | 383 ± 43 | 0.16 ± 0.02 |
| Heart | 5597 ± 138 | 783 ± 26 | 0.14 ± 0 |
| Muscle | 13 717 ± 508 | 1118 ± 101 | 0.08 ± 0 |

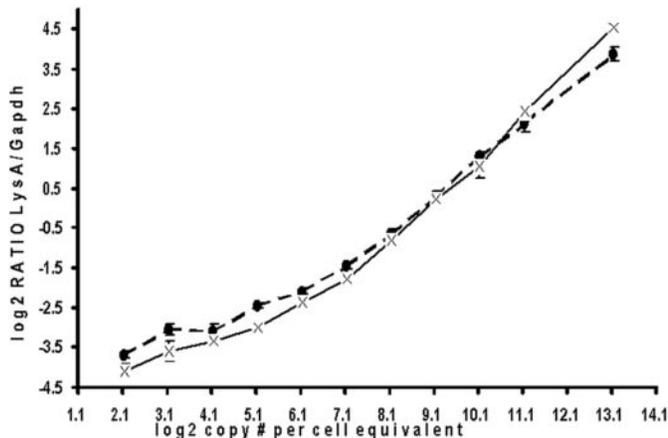


Figure 4. Standard curve for hybridization of increasing bacterial LysA gene versus constant Gapdh in liver cDNA mixtures (from Figure 3B). The ratio of the LysA fluorescence signal intensity versus the Gapdh signal intensity is plotted as the log 2 of LysA/Gapdh fluorescence intensities (y-axis). A 2-fold increase in the LysA abundance in the liver cDNA is plotted on the x-axis as the log 2 of the LysA copy number per cell equivalent. The log 2 values on the x-axis represent the following LysA copy numbers per liver cell equivalent of cDNA, log 2.1 = 4, 3.1 = 9, 4.1 = 18, 5.1 = 36, etc. cDNA mixtures were printed at two densities, either 400 pg total liver cDNA per spot (diamonds) or 800 pg liver cDNA per spot ('X'). An aliquot of 800 pg spots represent approximately 4000 cell equivalents of liver cDNA.

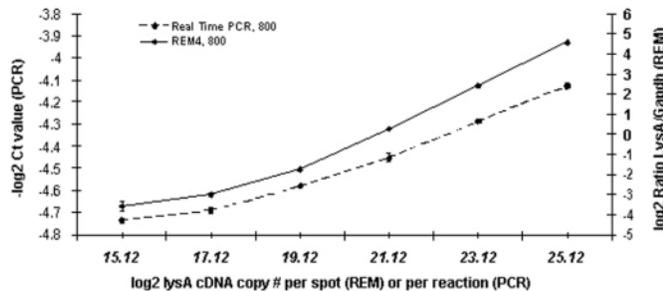


Figure 5. Comparison of REM technology with quantitative real-time PCR. REM data (diamonds); quantitative real-time PCR (circles). Left y-axis: log 2 of the δCt value for LysA concentration by real-time PCR. Right y-axis: log 2 of the ratio of LysA/Gapdh fluorescence intensities for standard mixtures from REM data (Figure 4). x-axis: log 2 of the LysA cDNA copy number per spot.

approximately four copies (log 2, 15.12) to 9100 copies (log 2, 25.12) of LysA per cell equivalent. Therefore, over a 2000-fold change in LysA abundance, REM technology is equal to quantitative real-time PCR in accuracy and sensitivity.

SMARTTM cDNAs used for REM production and gene expression analysis

A key feature of REM technology is its ability to represent a broad range of pathophysiological paradigms on a single product. However, valuable biological samples, such as biopsies or samples obtained by laser capture micro-dissection provide only small amounts of mRNA (29). This requires a method for amplification of the mRNA population while maintaining the relative balance in abundance between mRNA species (30–32). One approach is the production of ARNA (25).

Another approach is the cDNA amplification method termed as SMART (switching mechanism at the 5' end of the RNA template). The SMART method (22,24) utilized a combination of two primers in a single reverse transcription reaction. A tagged oligo(dT) primer is used to prime the first cDNA strand while the SMART oligonucleotide serves as a short, extended template at the 5' end of the RNA templates. When the reverse transcriptase reaches the 5' end of the mRNA, the enzyme switches templates and continues replicating to the end of the SMART oligonucleotide. PCR amplification is now initiated with primers complimentary to the 3' anchor and SMART oligonucleotide. This protocol uses a minimum number of PCR amplification cycles (~15) and SMART cDNAs have been shown to preserve the relative abundance of different mRNAs in complex cDNA mixtures (33–36).

We printed SMART cDNAs synthesized from mRNA isolated from five tumor/normal pairs from five major tumor types, including kidney, breast, uterus, lung and ovary. The SMART cDNAs were printed at 100, 200, 400 and 800 pg per spot, and each SMART cDNA sample was printed in quadruplicate. Hybridization of a REM-containing SMART cDNAs, with a single-stranded antisense Cy5-labeled probe to MYC, and a Cy3-labeled probe to beta 2 microglobulin (b2M), as a housekeeping reference, produced significant fluorescence signals across the whole range of printing densities (Figure 6A). The quadruplicate printing of each sample enabled us to calculate confidence intervals for each sample and draw a conclusion whether MYC was up- or down-regulated in the tumor from each tumor/normal pair. In the case of lung tumors, shown in Figure 6B, we concluded that MYC was up-regulated in all five tumors (100%). Up-regulation of MYC in the lung tumor samples was confirmed using quantitative real-time PCR. We calculated the $\delta\delta Ct$ value (37) for each tumor sample versus its matching normal sample (numbers above each tumor/normal pair in Figure 6B). A negative $\delta\delta Ct$ means that MYC was more abundant in the tumor sample compared to its matching normal sample.

The ratio for MYC expression versus b2M was calculated for all 200 SMART cDNA spots representing 25 tumor/normal pairs, quadruplicate spotted. This survey showed a predominant up-regulation of MYC in lung and ovary tumors and down-regulation in kidney tumors (Table 3). In contrast, however, MYC was predominantly unchanged in our group of breast and uterus tumors (Table 3). The highest MYC up-regulation was found in two lung tumors that had 4.3- and 5.7-fold increases, and the most significant down-regulation of MYC was in kidney tumors.

Different housekeeping genes and printing density yield similar results in REM technology

The choice of reference gene may be important in certain samples because routinely used housekeeping genes, such as Gapdh, are themselves regulated in certain instances. Therefore, in order to measure variability in REM data with different reference genes, we chose three different housekeeping genes, beta Actin, Ubiquitin and 23 kDa basic protein and hybridized them against a common test gene, glutathione peroxidase (GP), in three separate REMs. In addition, we prepared a mixture of the three reference probes and hybridized the mixture against

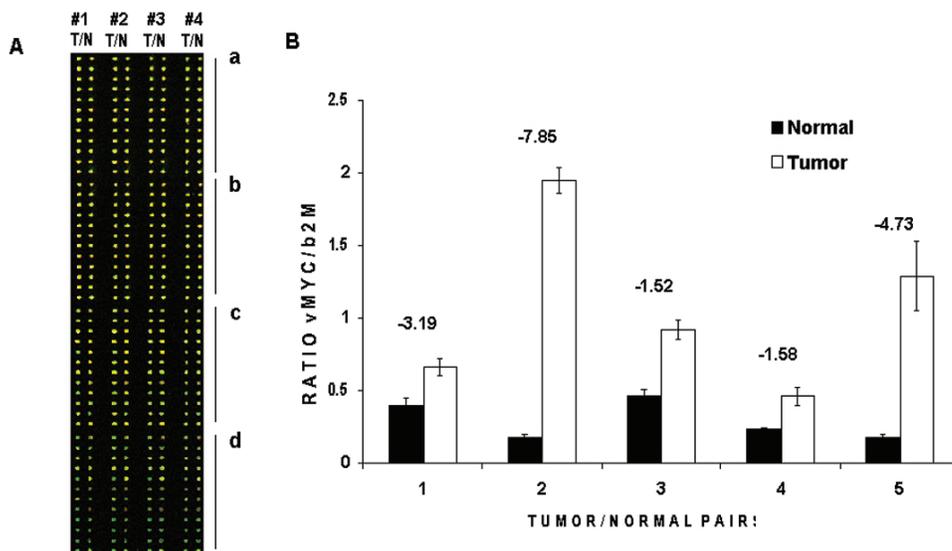


Figure 6. (A) Combined fluorescence image of a REM hybridized simultaneously with red (MYC) and green (b2M) probes. Image from segment of REM containing SMART cDNA samples of paired tumor/normal tissues is shown. Horizontal rows: images of four replicate sets of each tumor/normal pair (8 spots/row). Vertical columns: 12 tumor/normal pairs printed at either (a) 800 pg or (b) 400 pg or (c) 200 pg or (d) 100 pg SMART cDNA/spot. (B) Histogram of MYC expression in lung tumor and normal samples using b2M as the reference gene. Error bars represent standard deviation determined from four measurements of the cMYC/b2M ratio for each sample on the REM. The $\delta\delta CT$ values determined by real-time PCR are above each pair. A negative $\delta\delta CT$ value means MYC was higher in tumor tissue compared to companion normal tissue by quantitative real-time PCR analysis. $\delta\delta CT$ represents the difference in the number of real-time PCR cycles to reach maximum rate of amplification between tumor and normal paired samples.

Table 3. Quantitative analysis of MYC expression in a panel of 25 tumor/normal pairs

| Tissue | RAT N | SD | RAT T | SD | T/N | log ₂ T/N | Category | Summary |
|----------|-------|------|-------|------|-------|----------------------|-----------|---------------|
| Breast 1 | 1.03 | 0.16 | 0.66 | 0.05 | 0.64 | -0.64 | Down | |
| Breast 2 | 0.53 | 0.06 | 0.54 | 0.04 | 1.02 | 0.03 | Unchanged | 80% down |
| Breast 3 | 0.68 | 0.04 | 0.56 | 0.01 | 0.82 | -0.29 | Down | 20% unchanged |
| Breast 4 | 1.01 | 0.04 | 0.35 | 0.01 | 0.35 | -1.51 | Down | |
| Breast 5 | 0.55 | 0.12 | 0.28 | 0.04 | 0.51 | -0.97 | Down | |
| Kidney 1 | 0.96 | 0.04 | 0.21 | 0 | 0.22 | -2.18 | Down | |
| Kidney 2 | 1.21 | 0.09 | 0.43 | 0.05 | 0.36 | -1.47 | Down | 80% down |
| Kidney 3 | 1.21 | 0.08 | 0.3 | 0.05 | 0.25 | -2 | Down | 20% up |
| Kidney 4 | 1.03 | 0.11 | 0.33 | 0.05 | 0.32 | -1.64 | Down | |
| Kidney 5 | 0.73 | 0.07 | 1.34 | 0.09 | 1.84 | 0.88 | Up | |
| Lung 1 | 0.4 | 0.05 | 0.66 | 0.06 | 1.65 | 0.72 | Up | |
| Lung 2 | 0.18 | 0.02 | 1.95 | 0.09 | 10.83 | 3.44 | Up | 100% up |
| Lung 3 | 0.46 | 0.05 | 0.92 | 0.07 | 2 | 1 | Up | |
| Lung 4 | 0.24 | 0.01 | 0.46 | 0.06 | 1.92 | 0.94 | Up | |
| Lung 5 | 0.18 | 0.02 | 1.29 | 0.24 | 7.17 | 2.84 | Up | |
| Ovary 1 | 0.41 | 0.04 | 0.33 | 0.02 | 0.8 | -0.32 | Down | |
| Ovary 2 | 0.4 | 0.06 | 0.43 | 0.04 | 1.08 | 0.11 | Unchanged | 40% up |
| Ovary 3 | 0.33 | 0.03 | 1.36 | 0.02 | 4.12 | 2.04 | Up | 40% unchanged |
| Ovary 4 | 0.67 | 0.03 | 1.63 | 0.11 | 2.43 | 1.28 | Up | 20% down |
| Ovary 5 | 0.53 | 0.06 | 0.55 | 0.05 | 1.04 | 0.06 | Unchanged | |
| Uterus 1 | 0.41 | 0.06 | 0.37 | 0.04 | 0.9 | -0.15 | Unchanged | |
| Uterus 2 | 0.54 | 0.06 | 0.54 | 0.05 | 1 | 0 | Unchanged | 40% unchanged |
| Uterus 3 | 0.81 | 0.03 | 0.31 | 0.03 | 0.38 | -1.4 | Down | 40% down |
| Uterus 4 | 0.54 | 0.11 | 0.38 | 0.02 | 0.7 | -0.51 | Down | 20% up |
| Uterus 5 | 0.96 | 0.02 | 1.15 | 0.04 | 1.2 | 0.26 | Up | |

SMART cDNAs from five pairs each of lung, uterus, breast, ovary and kidney tumors are shown. Common reference gene was b2M. Data are from the 400 pg/spot series.

GP. The aim of this fourth REM was to determine whether the mixture of reference probes accurately reflected the results with each individual reference probe. All the probes were human genes and were hybridized to a REM containing the human tumor/normal pairs of SMART cDNAs. The data for

lung tumor/normal pairs from four REMs was representative of all the tumor types and is shown in Figure 7.

This analysis showed that GP was lower in lung tumor/normal pairs 1-4 and nearly the same in lung tumor/normal pair 5, in each of the four REMs. This confirmed that different

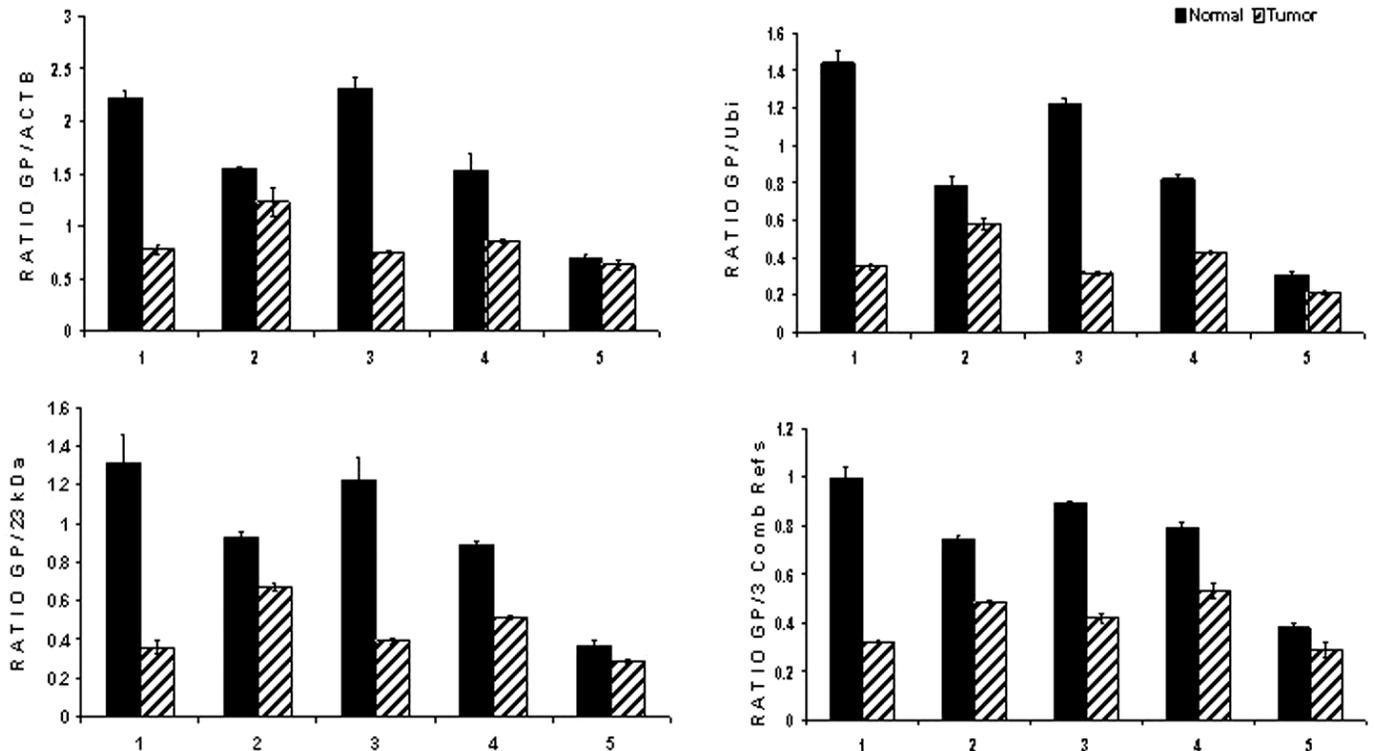


Figure 7. Common GP expression profiles obtained using three different reference genes and a mixture of the three reference genes. Data shown for five lung tumor/normal pairs from four separate REMs hybridized with the designated probes. Upper left, GP versus beta actin (ACTB); upper right, GP versus ubiquitin (Ubi); lower left, GP versus 23 kDa basic protein; and lower right, GP versus a mixture of all three reference genes. For each sample, the ratio of GP/reference signal is plotted. Standard errors are calculated from quadruplicate spotting of each sample.

reference genes can be used with qualitatively similar results. As expected the absolute ratios were different for each REM due to the different levels of hybridization of each housekeeping gene. However, the relative differences between tumor and normal samples were very similar across the set of four independently hybridized REMs. The REM hybridized with the combined set of three reference genes closely reflected the data from each reference gene singly. Therefore, the use of a combined reference probe may be preferable for REM hybridizations.

Finally, we compared the tumor/normal data across sets of identical samples that were printed at different printing densities. Data from a representative set of breast tumor/normal pairs printed at 200, 400 and 800 pg per spot and the average data for all three spotting densities are shown in Figure 8. Overall, there is a striking similarity between the datasets, supporting the conclusion that all three printing densities are suitable for REM analysis. In this example, GP was down-regulated in tumors in pairs 1 and 3, equal to normal in tumor/normal pairs 2 and 4 and very slightly up in tumor from pair 5. Each sample is quadruplicate spotted and therefore each bar in the average profile (D) represents 12 data points obtained at 3 densities for each sample. This type of multiple sampling is a unique strength of REMs that facilitates accurate standard error measurements and the detection of small differences between tumor and normal samples.

DISCUSSION

In this report, we have validated REM technology for measuring the expression of test genes in a diverse spectrum of biological samples in a high-throughput manner. We have demonstrated that the REM technology can detect organ preferential gene expression of both abundant transcripts such as albumin in the liver, and rare transcripts such as hepatocyte nuclear factor 4 (Hnf4) in the liver and insulin-like binding protein 1 (Igf1), expression in pancreas and spleen (Table 1). We also detected MYC oncogene expression in both tumor and normal human tissue samples and confirmed the differential regulation with quantitative real-time PCR.

One feature of REM technology is that it can be used to measure gene expression differences that are due to sex of the individual. For example, in a prototype REM, we included liver cDNAs from male and female mice that were either C57B16 or CD1 genetic backgrounds. Using the REM, we showed that albumin expression was not different in livers of mice from different genetic backgrounds; however, the sex of the mouse had a significant effect on albumin expression (Figure 2). The data showed that albumin expression is significantly higher in female liver. This is not generally appreciated, however, it is consistent with the biological functions of albumin which include being the major serum binding protein for the female hormone, estrogen (27). Second generation REMs, which contain 5–10 replicate organ samples, from both male and female mice of different genetic

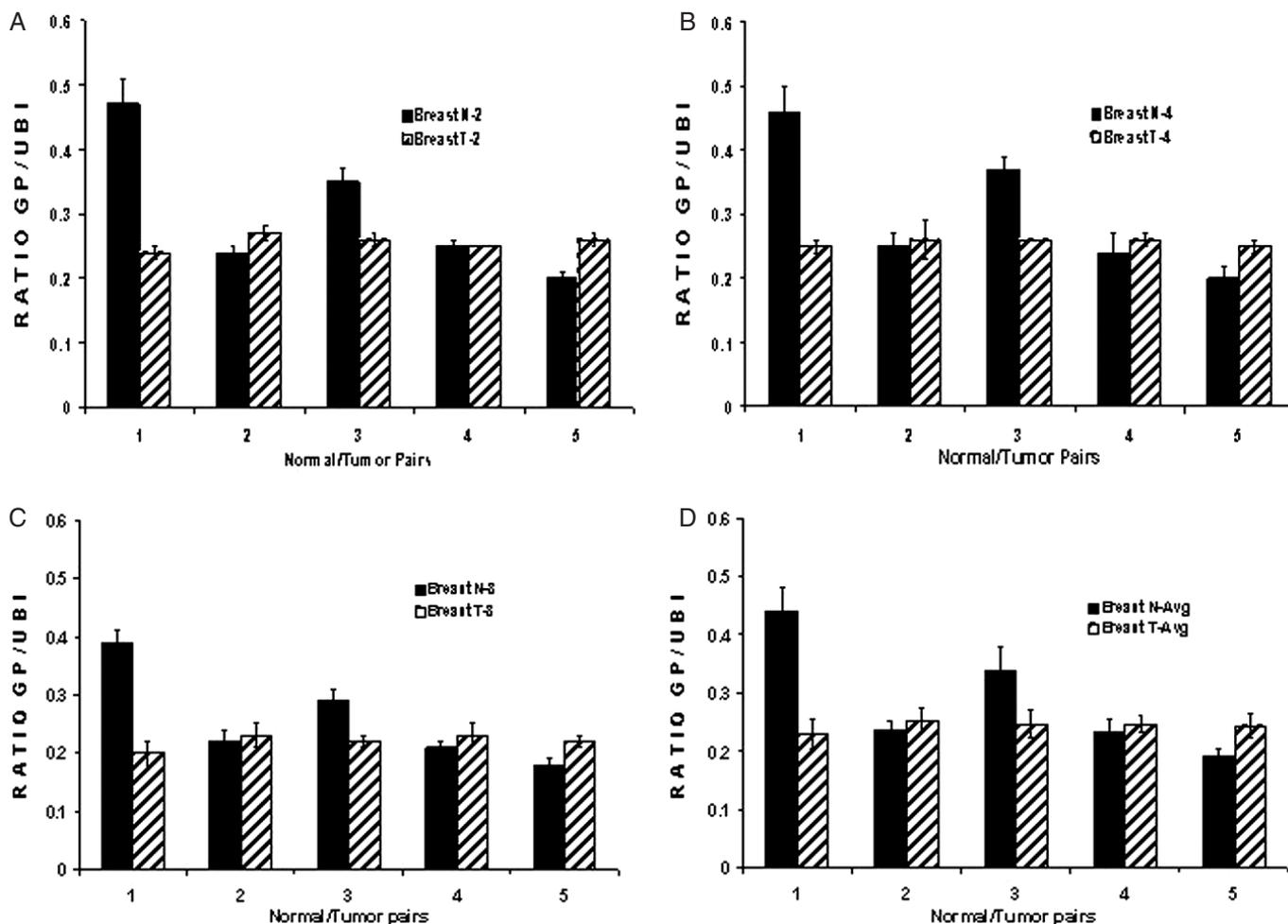


Figure 8. Common expression profiles obtained for tumor/normal samples printed at different densities. Data shown for five breast tumor/normal pairs hybridized with GP and ubiquitin (UBI). SMART cDNAs were printed at three printing densities. (A) An aliquot of 200 pg of SMART cDNA per spot; (B) 400 pg of SMART cDNA per spot; (C) 800 pg of SMART cDNA per spot; and (D) average data from all three levels. The REM containing human tumor/normal pairs was hybridized with human GP probe plus human ubiquitin probe. For each sample, the ratio of GP/ubiquitin signal is plotted. Standard errors are calculated from quadruplicate spotting of each sample. Data shown are for five kidney tumor/normal pairs.

backgrounds will have the unique ability to detect previously unappreciated gene expression differences due to sex and genotype.

Any microarray technology that utilizes printing of nucleic acids must have a means to control for variability in printing density that invariably occurs between samples. The need for controls for differential loading are one of the main limitations of earlier RNA and cDNA dot blots (19). Using nylon arrays, it has been necessary to elute a first probe and rehybridize the array with a second housekeeping gene in order to normalize the signals for the first probe. The use of a glass printing format and co-hybridization with two fluorescent dyes has eliminated the need for re-hybridization for REM technology.

Two lines of evidence in this report support the conclusion that REMs can be accurately normalized by co-hybridization with a housekeeping gene. First, we showed that cDNAs from muscle and other organs, which do not express albumin, when normalized against Gapdh hybridization, reveal a very low ratio of albumin/Gapdh that is essentially equivalent to background (Table 2). We showed that this internal normalization

function well for samples that were apparently 10 or more fold different in loading per spot (Table 2). Second, when we purposely loaded different amounts of the same samples on a REM, we repeatedly observed sets of ratios of test gene to normalization gene that were virtually identical across cDNA spotting densities from 200 to 800 pg/spot (Figure 8). The same degree of reproducibility and sensitivity of gene expression differences detected through this multiple spotting approach has not been previously reported for microarrays.

In a pilot experiment, we tested whether REMs could be eluted and re-hybridized with two new probes. We successfully eluted the samples and re-hybridization yielded significant signals that generally had similar ratios of test to reference genes. However, the signals were reduced in intensity compared to the first hybridization (data not shown). Therefore, the pilot data strongly suggest that conditions will be found that will enable the re-hybridization of REMs, thus greatly extending their usefulness.

Another possible expanded use of REMs includes the use of more than two labeled probes per hybridization. Theoretically,

the number of probes that could be hybridized simultaneously will only be limited to the number of fluorescence signals that can be distinguished by a laser detector. Therefore, it may be possible to simultaneously hybridize REMs with sets of probes that detect as many as 5–10 genes in a particular pathway, simultaneously.

The sensitivity and accuracy of REM technology was demonstrated by the use of artificial mixtures. Data from a prototype REM that contained standardized mixtures of a bacterial gene, *LysA* into a liver cDNA showed that a specific hybridization signal was detected when as few as four copies of the *LysA* cDNA were present per liver cell cDNA equivalent. Furthermore, the *LysA*/*Gapdh* ratio increased in a near linear pattern until 9000 copies per cell was reached. This relationship was repeated whether we printed the liver cDNAs at 400 or 800 pg/spot demonstrating again the wide range of the experimental sample loading for which accurate measurements can be obtained (Figure 4).

The accuracy of the standard curve data was tested by analysis of the standard mixtures with quantitative real-time PCR. In this analysis, quantitative real-time PCR and REM data closely paralleled each other from samples with 4 to 9000 copies of *LysA* per liver cell cDNA equivalent. Therefore, we conclude that REM technology is equivalent to quantitative real-time PCR over at least a 2000 to 4000-fold change of mRNA concentrations.

The SMART method has been successfully applied to the generation of full-length cDNA libraries (24), and as a source for cDNA probes for GEMS from RNA obtained by laser capture microdissection (32). When we printed SMART DNAs on a REM at multiple printing densities of spotting, the ratio of expression of test genes were nearly identical in the range of spotting densities (Figure 7).

There has been much discussion in the literature about the preservation of original RNA representation after mRNA/cDNA amplification. It is generally assumed that linear amplification is superior to exponential amplification due to biases in abundance relationships (38,39). However, this assumption does not hold up on closer examination of the recent literature. Wang *et al.* (40) has pointed out that conventional T₇-based RNA amplification can introduce biases in the amplification because of a possible 5'-under representation and because low stringency temperatures are applied during generation of the double-stranded cDNA. SMART cDNA amplification from total RNA was found to preserve representation of high, medium and low abundance mRNAs and compared favorably to quantitative northern-blot analysis (41). Additionally, SMART cDNA generated signals expressed nearly identical patterns to unamplified total RNA probes upon hybridization to 4600 arrayed genes in a GEM analysis (41).

It is known that genes that are generally considered to be housekeeping genes, such as *Gapdh*, are differentially expressed under various experimental conditions. Therefore, it would be advantageous to be able to utilize a mixture of housekeeping genes for normalization, in order to control for minor variations in any one of the genes. In this report, we have utilized four different normalization genes including *Gapdh*, beta Actin, Ubiquitin and 23 kDa basic protein. In all the cases, where the same test gene was tested against two or more reference genes, the data were qualitatively similar (Figure 7). Furthermore, data from REMs hybridized with

a mixture of reference genes produced datasets that were virtually indistinguishable from those of REMs hybridized with only a single reference gene (Figure 7). Therefore, the use of standard mixtures of reference genes can control for both loading differences and small variations in expression of housekeeping genes.

We investigated the expression of the oncogene, *MYC* in the SMARTTMcDNA tumor/normal pairs. Data from quadruplicate spotted samples enabled us to calculate confidence intervals for the ratios for each tumor and normal sample. Therefore, we were able to draw conclusions as to differential oncogene expression in the paired samples at a level of sensitivity not previously possible for RNA dot blots (19). In a single REM containing 25 tumor/normal pairs from five tumor types, we were able to calculate the relative *MYC* expression and reach conclusions as to whether *MYC* was up, down or unchanged in the whole panel of tumor samples (Table 3). This survey serves as one example of how REM technology can be used in cancer research. Since thousands of samples can be printed on a single REM, REM technology can provide a high-throughput approach to testing candidate gene expression in diverse tissues and tumors.

The only other array based approach designed to test samples from multiple tissue types simultaneously is tissue microarrays (20). These arrays contain thin sections of multiple tissues on a microscope slide, allowing an investigator to determine gene expression using antibodies to detect protein in cells. Also, *in situ* hybridization of tissue microarrays provides information on the expression of a gene in specific cell types. However, antibody staining and *in situ* hybridization are not quantitative technologies and are labor intensive. Furthermore, due to the nature of the experimental approach each section cut from the tissue microarray is different from the previous section. Also, these arrays are generally limited to fewer than one hundred samples, whereas REMs can easily accommodate thousands of samples that are spotted more than once for quantitative analysis.

Therefore, REM Technology fulfills an important need for a high-throughput, sensitive, accurate and quantitative method to measure gene expression simultaneously in multiple tissues or cell types, at a time in biological research when there is a strong emphasis on quantitative expression analysis. REMs provide a platform on which to build libraries of samples that can be used to characterize specific functions of genes in specific biological contexts. In addition to general survey REMs that contain samples from organs, tissues and cell types, specialty REMs that have experimental samples designed to ask specific questions about regulation of a gene in specific cellular contexts, and developmental contexts, can be designed and produced. The future content and number of specialty REMs (such as liver, kidney, heart, tumor profiles, developmental stages and gene knockout REMs) and their application to biology is virtually unlimited. Therefore, we envision the library of REMs as continuing to grow and the impact of REMs on biological research to increase with time. The availability of REMs that have samples from classic experiments will provide researchers with access to relate their current research directly to historically validated paradigms. The use of reference REMs designed to ask questions in the area of toxicology and pharmaceutical research may also gain use in the drug approval process. REMs can essentially

'immortalize' specific experiments that can be printed thousands of times and be widely distributed.

ACKNOWLEDGEMENTS

The authors thank Dr Liang Zhu and Dr David Shafritz for critical reading of this manuscript. This work was supported by funding from the National Institute of Health Grants RO1CA37232 to C.E.R., Liver Center grant (5P30DK41296) and Einstein Comprehensive Cancer Center grant (CA13330-32).

REFERENCES

- Brown,P.O. and Botstein,D. (1999) Exploring the new world of the genome with DNA microarrays. *Nature Genet.* (suppl.), **21**, 33–37.
- Lewin,B. (1997) Eukaryotic gene expression, chapters 28 and 29. In B. Lewin (ed.), *Genes VI*. Oxford University Press, Oxford, NY and Tokyo.
- Watson,J.D., Hopkins,N.H., Roberts,J.W., Steitz,J.A. and Weiner,A.M. (1987) The steps in protein synthesis, chapters 13–15. In Gillen,J.R. (ed.), *Molecular Biology of the Gene*. Benjamin/Cummings Publishing Co., Menlo Park, CA.
- Shalon,T.D. (1995) DNA microarrays: a new tool for genetic analysis. PhD Thesis, Stanford University, UMI Dissertation services, Ann Arbor, MI, pp. 29–33.
- 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.
- Schena,M., Shalon,D., Heller,R., Chai,A., Brown,P.O. and Davis,R.W. (1996) Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. *Proc. Natl Acad. Sci. USA*, **93**, 10614–10619.
- Iyer,V.R., Eisen,M.B., Ross,D.T., Schuler,G., Moore,T., Lee,J.C.F., Trent,J.M., Staudt,L.M., Hudson,J., Jr, Boguski,M.S., Lashkari,D., Shalon,D., Botstein,D. and Brown,P.O. (1999) The transcriptional program in the response of human fibroblasts to serum. *Science*, **283**, 83–87.
- Robert,R., Klevecz,J.B., Forrest,G. and Murray,D.B.A. (2004) Genome wide oscillation in transcription gates DNA replication and cell cycle. *Proc. Natl Acad. Sci. USA*, **101**, 1200–1205.
- Plescia,C.P., Rogler,C.E. and Rogler,L.E. (2001) Genomic expression analysis implicates Wnt signaling pathway and extracellular matrix alterations in hepatic specification and differentiation of murine hepatic stem cells. *Differentiation*, **68**, 254–269.
- Yu,Y., Khan,J., Khanna,C., Helman,L., Meltzer,P.S. and Merlino,G. (2004) Expression profiling identifies the cytoskeletal organizer ezrin and the developmental homeoprotein Six-1, as key metastatic regulators. *Nature Med.*, **10**, 175–181.
- Mathiassen,S., Lauemoller,S.L., Ruhwald,M., Claesson,M.H. and Buus,S. (2001) Tumor-associated antigens identified by mRNA expression profiling induce protective anti-tumor immunity. *Eur. J. Immunol.*, **31**, 1239–1246.
- Kaminski,N., Allard,J.D., Pittet,J.F., Zuo,F., Griffiths,M.J.D., Morris,D., Huang,Z., Sheppard,D. and Heller,R.A. (2000) Global analysis of gene expression in pulmonary fibrosis reveals distinct programs regulating lung inflammation and fibrosis. *Proc. Natl Acad. Sci. USA*, **97**, 1778–1783.
- Perou,C.M., Jeffrey,S.S., van de Rijn,M., Rees,C.A., Eisen,M.B., Ross,D.T., Pergamenschikov,A., Williams,C.F., Zhu,S.X., Lee,J.C. *et al.* (1999) Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. *Proc. Natl Acad. Sci. USA*, **96**, 9212–9217.
- Perou,C.M., Sorlie,T., Eisen,M.B., van de Rijn,M., Jeffrey,S.S., Rees,C.A., Pollack,J.R., Ross,D.T., Johnsen,H., Akslen,L.A. *et al.* (2000) Molecular portraits of human breast tumours. *Nature*, **406**, 747–752.
- Allzaeh,A.A., Eisen,M.B., Davis,R.E., Ma,C., Lossos,I.S., Rosenwald,A., Boldrick,J.C., Sabet,H., Tran,T., Yu,X. *et al.* (2000) Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature*, **403**, 503–511.
- Takahashi,M., Rhodes,D.R., Furge,K.A., Kanayama,H., Kagawa,S., Haab,B.B. and Teh,B.T. (2001) Gene expression profiling of clear cell renal cell carcinoma: gene identification and prognostic classification. *Proc. Natl Acad. Sci. USA*, **98**, 9754–9759.
- Bittner,M., Meltzer,P., Chen,Y., Jiang,Y., Seftor,E., Hendrix,M., Radmacher,M., Simon,R., Yakhini,Z., Ben-Dor,A. *et al.* (2000) Molecular classification of cutaneous malignant melanoma by gene expression profiling. *Nature*, **406**, 536–540.
- Kitahara,O., Furukawa,Y., Tanaka,T., Kihara,C., Ono,K., Yanagawa,R., Nita,M.E., Takagi,T., Nakamura,Y. and Tsunoda,T. (2001) Alterations of gene expression during colorectal carcinogenesis revealed by cDNA microarrays after laser-capture microdissection of tumor tissues and normal epithelia. *Cancer Res.*, **61**, 3544–3549.
- Kafatos,F.C., Jones,C.W. and Efstratiadis,A. (1979) Determination of nucleic acid sequence homologies and relative concentrations by a dot hybridization procedure. *Nucleic Acids Res.*, **7**, 1541–1552.
- Kononen,J., Bubendorf,L., Kallioniemi,A., Bärklund,M., Schraml,P., Leighton,S., Torhorst,J., Mihatsch,M.G., Sauter,G. and Kallioniemi,O.P. (1998) Tissue microarrays for highthroughput molecular profiling of tumor specimens *Nature Med.*, **4**, 844–847.
- Sambrook,J. and Russell,D.W. (2001) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Chenchik,A., Zhu,Y.Y., Diatchenko,L., Li,R., Hill,J. and Siebert,P.D. (1998) Generation of high quality cDNA from small amounts of total RNA by SMART PCR. In Siebert,P.D. and Larrick,J.W. (eds), *Gene Cloning and Analysis of RT-PCR*. Eaton Publishing, Natick, MA, pp. 305–319.
- Zumabayeva,B., Diatchenko,L., Chenchik,A. and Siebert,P.D. (2001) Use of SMART generated cDNA for gene expression studies in multiple human tumors. *BioTechniques*, **30**, 1–7.
- Zhu,Y.Y., Machleder,E.M., Chenchik,A., Li,R. and Siebert,P.D. (2001) Reverse transcriptase template switching: a SMART™ approach for full-length cDNA library construction. *BioTechniques*, **30**, 892–897.
- Eberwine,J. (1996) Amplification of mRNA populations using aRNA generated from immobilized oligo(dT)-T7 primed cDNA. *BioTechniques*, **20**, 584–591.
- Cheung,V.G., Morley,M., Aguilar,F., Massimi,A., Kucherlapati,R. and Childs,G. (1999) Making and reading microarrays. *Nature Genet.*, **21**, 15–19.
- Andre,C., Jacquot,Y., Truong,T.T., Thomassin,M., Robert,J.F. and Guillaume,Y.C. (2003) Analysis of the progesterone displacement of its human serum albumin binding site by beta-estradiol using biochromatographic approaches: effect of two salt modifiers. *J. Chromatogr. B*, **769**, 267–281.
- Winer,J., Kwang,C., Jung,S., Shackel,I. and Williams,P.M. (1999) Development and validation of real-time quantitative reverse transcriptase–polymerase chain reaction for monitoring gene expression in cardiac anal. *Biochemistry*, **270**, 41–49.
- Emmert-Buck,M.R., Bonner,R.F., Smith,P.D., Chuaqui,R.F., Zhuang,Z., Goldstein,S.R., Weiss,R.A. and Liotta,L.A. (1996) Laser capture microdissection. *Science*, **27**, 467–1001.
- Luo,L., Salunga,R.C., Guo,H., Bittner,A.K., Joy,C., Galindo,J.E., Xias,H., Rogers,K.E., Wan,J.S., Jackson,M.R. *et al.* (1999) Gene expression profiles of laser-captured adjacent neuronal subtypes. *Nature Med.*, **5**, 117–121.
- Luzzi,V., Mahadevappa,M., Raja,R., Warrington,J.A. and Watson,M.A. (2003) Accurate and reproducible gene expression profiles from laser capture microdissection, transcript amplification, and high density oligonucleotide microarray analysis. *J. Mol. Diagn.*, **5**, 9–14.
- Fink,L., Kohlhoff,S., Stein,M.M., Hanzel,J., Weissmann,N., Rose,F., Akkayagil,E., Manz,D., Grimminger,F., Seeger,W. and Bohle,R.M. (2002) cDNA array hybridization after laser-assisted microdissection from nonneoplastic tissue. *Am. J. Pathol.*, **160**, 81–90.
- Wang,E., Miller,L.D., Galen,A., Ohnmacht,G.A., Liu,E.T. and Marinocola,F.M. (2000) High-fidelity mRNA amplification for gene profiling. *Nat. Biotechnol.*, **18**, 457–459.
- Kestler,D.P., Hill,M., Agarwal,S. and Hall,R.E. (2000) Use of bidirectional blots in differential display analysis. *Anal. Biochem.*, **280**, 216–220.

35. Becker, B., Vogt, T., Landthaler, M. and Stolz, W. (2001) Detection of differentially regulated genes in keratinocytes by cDNA array hybridization: Hsp27 and other novel players in response to artificial ultraviolet radiation. *J. Invest. Dermatol.*, **116**, 983–988.
36. Vernon, S.D., Unger, E.R., Rajeevan, M., Dimulescu, I.M., Nisenbaum, R. and Campbell, C.E. (2000) Reproducibility of alternative probe synthesis approaches for gene expression profiling with arrays. *J. Mol. Diagn.*, **2**, 124–127.
37. Livak, K. and Schmittgen, T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2- $\Delta\Delta C_t$ method. *Methods*, **25**, 402–408.
38. Baugh, L.R., Hill, A.A., Brown, E.L. and Hunter, C.P. (2001) Quantitative analysis of mRNA amplification by *in vitro* transcription. *Nucleic Acids Res.*, **29**, 1–9.
39. Dixon, A.K., Richardson, P.J., Pinnock, R.D. and Lee, K. (2000) Gene-expression analysis at the single-cell level. *Trends Pharmacol.*, **21**, 65–70.
40. Wang, E., Miller, L.D., Ohnmacht, G.A., Liu, E.T. and Marincola, F.M. (2000) High-fidelity mRNA amplification for gene profiling. *Nat. Biotechnol.*, **18**, 457–459.
41. Endege, W.O., Steinmann, K.E., Boardman, L.A., Thibodeau, S.N. and Schlegel, R. (1999) Representative cDNA libraries and their utility in gene expression profiling. *BioTechniques*, **26**, 542–550.