

Detection of Differentially Expressed Genes in an Isogenic Breast Metastasis Model using RNA Arbitrarily Primed-Polymerase Chain Reaction Coupled with Array Hybridization (RAP-Array)

Derek D. Sloan,* Ben Nicholson,*
Virginia Urquidi,[†] and Steve Goodison[†]

From the University of California San Diego,* La Jolla, California; and the Department of Pathology and Cancer Center,[†] University of California San Diego, La Jolla, California

To facilitate the study of the mechanisms of breast cancer metastasis we have previously characterized a pair of breast tumor cell lines that originate from the same breast tumor cell line MDA-MB-435, but which have diametrically opposite metastatic capabilities. These cell lines constitute a stable and accessible experimental system for the identification of metastasis-related genes and for the study of their role in the process of metastasis. In this study, we used a combination of RNA arbitrarily primed-polymerase chain reaction (RAP-PCR) fingerprinting and cDNA arrays (here termed "RAP-array") to identify genes differentially expressed with respect to metastatic phenotype. RAP-PCR was used to generate radioactive probes of reduced complexity for hybridization to nylon membranes containing 588 cDNAs of known identity. Single RAP-PCR fingerprint probes hybridized from 61 (10.4%) to 116 (19.7%) of the filter array targets, with a signal detection overlap of ~21%. A total of 344 (57%) of the 588 target genes were detected by five single RAP-PCR fingerprints. The advantage of using reduced complexity probes was highlighted by the fact that the combination of RAP probes before hybridization compromised the overall detection rate by up to 40%. Sequential application of RAP-PCR probes allowed the screening of a greater, and an alternative fraction of the transcript population than was achieved with a radiolabeled total cDNA probe. Verification by quantitative reverse transcriptase-PCR confirmed significantly increased expression of keratin 9 (>100-fold) in nonmetastatic breast tumor cells and of CD70 (fivefold) in metastatic cells. The differential expression of keratin 9 and CD70 was maintained between cells grown as primary xenografts in athymic mice. The RAP-array method enabled the detection of genes not revealed using other screening methods and that are candidates for further investi-

gation in the context of metastatic phenotype. (*Am J Pathol* 2004, 164:315-323)

The multistep nature of metastasis poses difficulties in both design and interpretation of experiments to unveil the mechanisms causing the process. To facilitate such studies we have recently developed and characterized an isogenic cell line model.¹ By serial dilution cloning of the MDA-MB-435 breast tumor cell line and screening by orthotopic implantation into the mammary fat pad of athymic mice, a pair of breast tumor cell lines (M-4A4 and NM-2C5) that originate from the same breast tumor, but that have diametrically opposite metastatic capabilities were derived. In 74% of inoculated athymic mice clone M-4A4 metastasized consistently to the lungs and the lymph nodes, mimicking major dissemination routes of human breast cancer. Conversely, although equally tumorigenic, clone NM-2C5 did not metastasize to any distal site. We have confirmed that the cell lines originate from a single genetic source by spectral karyotyping and evaluated the expression of a number of proteins previously implicated in cellular transformation and metastasis. Initial targeted and exploratory RNA and protein analyses have revealed that the secreted proteins thrombospondin-1 and osteopontin are differentially expressed.¹ Subsequently, representational difference analysis of cDNA obtained from the two clonal populations revealed a correlation between the increased expression of tyrosinase-related protein-1 and the matrix metalloproteinase-8 genes with the nonmetastatic phenotype.² These cell lines constitute a stable and accessible model for the identification of genes involved in the multistep process of breast tumor metastasis.

Modern technological advances now permit the application of high-throughput gene expression analysis to identify patterns of expression involved in complex biological processes. While the field of gene expression

Supported in part by a Sidney Kimmel Scholar Award (to S. G.) and the National Institutes of Health (grant UO1 CA84998).

Accepted for publication October 2, 2003.

Address reprint requests to Steve Goodison Ph.D., Department of Pathology, University of Florida, 655 W. 8th Street, Jacksonville, FL 32209-6511. E-mail: stevegoodison@jax.ufl.edu.

profiling continues to evolve, the majority of current methods use cDNA or oligonucleotide targets attached to a solid support. Hybridization using labeled sample cDNA populations as a probe enables the expression profile of many genes to be simultaneously determined. However, labeling the total cDNA derived from a given cellular source results in an extremely complex mixture representing literally tens of thousands of transcripts. This probe complexity severely compromises the ability to detect the rarer mRNA species.³ It is reasonable to state that current expression array approaches are biased toward transcript abundance. Although advances in the use of fluorescent labeling and detection systems have increased the sensitivity^{4,5} and the dynamic range of detection, to achieve the accurate monitoring of relatively rare transcripts, some reduction in the probe complexity is necessary.

A probe having reduced complexity and, therefore, increased representation of rare messages can be created using RNA arbitrarily primed-polymerase chain reaction (RAP-PCR) coupled with array hybridization fingerprinting.⁶ RAP-PCR samples a reproducible subset of the message population based on selective amplification with arbitrary primers.^{7,8} A typical RAP-PCR fingerprint contains between 1000 to 2000 DNA fragments, including products synthesized from relatively rare mRNAs, resulting from an appropriate match with the arbitrarily designed amplification primers. A single probe derived from RAP-PCR can detect in the order of 10 to 20% of targets on an array, a considerable improvement over the visualization of fingerprints possible on denaturing polyacrylamide gels. Furthermore, repeated application of distinct RAP-PCR probes allows a greater fraction of the message population to be screened on this type of array than can be achieved with a radiolabeled total cDNA probe.⁶ The application of RAP-PCR probes to arrays avoids the need for gel purification and sequencing necessary with conventional differential display methodology because sequence information of the array targets is immediately available.

In this report, we show that the detection of differential expression in our breast metastasis model system can be greatly facilitated using RAP-array analysis and that this method allowed us to identify genes that we had not found previously using other screening methods. The application of five RAP-PCR probes resulted in hybridization to 334 individual array targets (57% coverage), from which 24 putative differentially expressed genes were identified. Differential expression of several of these candidates was confirmed using quantitative reverse transcriptase-PCR (qPCR) analysis. Keratin 9, TIMP3, and CTGF transcripts were found to be significantly up-regulated in the nonmetastatic NM-2C5 cell line, whereas CD70, CLK3, erbB4, and CDK2 transcripts were more highly expressed in metastatic M-4A4 cells. Of these candidates, three genes (*keratin 9*, *CTGF*, and *CD70*) were proven to maintain specific differential expression *in vivo*, as evidenced by qPCR analysis of RNA recovered from primary tumor xenografts formed in athymic mice. To obtain the greatest possible coverage of the transcriptome for profiling purposes, a broad analytical approach

is required. RAP-array analysis constitutes an effective approach for the detection and identification of low-abundance transcripts. Subsequent genetic manipulation of the candidate genes identified with this method in our clonal cell lines will enable the evaluation of their functional significance in the dissemination of breast cancer.

Materials and Methods

Tissue Culture and RNA Preparation

The monoclonal M-4A4 and NM-2C5 cell lines were propagated in RPMI 1640 supplemented with 10% newborn calf serum (Life Technologies, Inc., Gaithersburg, MD) at 37°C in a humidified atmosphere of 5% CO₂ and 95% air. Cells were harvested at ~75% confluency by direct application of RNA-easy lysis buffer (Qiagen, Valencia, CA) and homogenized by needle-shearing. Frozen xenograft tissues were mechanically homogenized in chaotropic buffer. The RNA-easy purification kit (Qiagen) was used to isolate total RNA. For direct cDNA probe production, poly A⁺ mRNA was isolated from total RNA using poly dT-coated latex beads (Invitrogen, San Diego, CA). The RNA was incubated with 0.08 U/ μ l of RNase-free DNase (Promega, Madison, WI) for 40 minutes at 37°C. RNA quantity was measured by spectrophotometry, checked for quality by agarose gel electrophoresis, and stored at -80°C.

Preparation of cDNA Probes

One hundred ng and 1 μ g of polyA⁺ RNA was reverse-transcribed with the MMLV enzyme (Promega) and the appropriate cDNA synthesis (CDS) primer mix (Clontech, Palo Alto, CA) according to the Atlas Array user's manual. This primer mix is supplied with the Atlas cDNA Expression Array and ensures that cDNAs are only synthesized for the genes on a particular Atlas Array. The reverse transcriptase reaction (50°C/25 minutes) was performed in the presence of 35 μ Ci of [α -³²P]dATP (3000 Ci/mmol; ICN, Irvine, CA). Labeled cDNA was purified with the Qiagen Nucleotide Removal kit (Qiagen) and radioactive nucleotide incorporation was monitored by scintillation counting.

RNA Fingerprinting

RAP-PCR was performed according to published protocols.⁹ Reverse transcription of total RNA was performed at 37°C for 1 hour with an oligo (dT) primer and MMLV-reverse transcriptase (Promega), using two concentrations of RNA from each sample (5 μ g and 500 ng RNA per cDNA reaction). PCR was performed after the addition of a pair of 10-mer oligonucleotide primers of arbitrary sequence; pair A: 5'-GCCACCCAGA and 5'-GTAGCCCAGC; pair B: 5'-ACGAAGAAGC and 5'-AGGGCACCAC; pair C: 5'-CC-AGTGGAGG and 5'-AGTGAGCACG; pair D: 5'-TGGAAC-CGAG and 5'-ATCGTGCTGG; pair E: 5'-TCAGGGCTCC and 5'-GGCAAGCGTC. cDNA (5 μ l of 20 μ l reaction) was subjected to PCR in a mixture containing 20 mmol/L Tris, pH

8.3, 20 mmol/L KCl, 6.25 mmol/L MgCl₂, 0.2 mmol/L of each dNTP (except dCTP at 0.07 mmol/L), 2 μ Ci [α -³²P]dCTP (3000 Ci/mmol, ICN), 2 μ mol/L of each oligonucleotide primer, and 5 U Amplitaq DNA polymerase Stoffel fragment, (Perkin-Elmer-Cetus, Norwalk, CT) in a 50- μ l final reaction. The PCR reaction was performed using 35 cycles of 94°C for 1 minute, 35°C for 1 minute, and 72°C for 2 minutes. Unincorporated nucleotides and primers were removed from the reaction with the Qiagen Nucleotide Removal kit (Qiagen) and radioactive nucleotide incorporation monitored by scintillation counting. The RNA fingerprint was checked by visualization via gel electrophoresis before use in hybridizations. An aliquot of the amplification reaction (2.5 μ l) was mixed with 7.5 μ l of formamide dye solution, denatured at 85°C for 4 minutes, and chilled on ice. The aliquot (2.5 μ l) was loaded onto a 6% polyacrylamide/1 \times Tris borate-ethylenediaminetetraacetic acid gel. Gels were either exposed to phosphorimaging screens or stained with SYBR Green I dye (Sigma-Aldrich) and photographed using the Bio-Rad Chemi-Doc system (Bio-Rad, Hercules, CA).

Array Hybridization

The Atlas Human Cancer cDNA array (Clontech) contains 588 known cDNA targets spotted on positively charged nylon membranes. Plasmid and bacteriophage DNAs are included as negative controls to confirm hybridization specificity and several housekeeping gene cDNAs are included as positive controls and for normalization of data between arrays. Membranes were prehybridized by incubation in a roller bottle at 68°C for 30 minutes with ExpressHyb solution (Clontech) containing 100 μ g/ml of fragmented, denatured salmon-sperm DNA (Sigma). Labeled probe was denatured by boiling in the presence of Cot-1 DNA (Life Technologies, Inc.), followed by incubation on ice. The prehybridization solution was exchanged with 5 ml of prewarmed (68°C) ExpressHyb hybridization solution containing 10 to 20 \times 10⁶ cpm of labeled probe, 100 μ g/ml of fragmented, denatured salmon sperm DNA, and 10 ng/ml of poly (dA) to block oligo (dT) stretches in the radiolabeled probe. Hybridization continued at 68°C for 18 hours in a roller bottle. Primary washes (3 times/30 minutes) were performed with continuous agitation at 68°C with prewarmed 2 \times standard saline citrate and 1% sodium dodecyl sulfate. A secondary wash at 68°C/30 minutes was performed with 0.1 \times standard saline citrate, 0.5% sodium dodecyl sulfate. One final 10-minute wash was performed at room temperature with 2 \times standard saline citrate only. Filters were blotted, wrapped in Saran wrap while moist, and imaged using a Storm Phosphorimager (Amersham Biosciences, Piscataway, NJ). Filters were reused up to three times. Stripping of probe was achieved by boiling in 0.5% sodium dodecyl sulfate for 5 to 10 minutes. After cooling in the solution for 15 minutes the filter was rinsed with 2 \times standard saline citrate and 1% sodium dodecyl sulfate, immediately wrapped in Saran wrap to maintain moisture, and exposed to phosphorimaging screens to check for successful stripping of radioactive probe. For probe combinations, the specific activity (10 \times 10⁶ cpm) of each probe was maintained so

as to be comparable to that used in the single probe hybridization experiment. The resulting hybridization solution (up to 50 \times 10⁶ cpm of labeled probe) was added to the filters and processed as described above.

Image and Data Analysis

Phosphorimaging was performed using a Molecular Dynamics Storm system (Amersham). Images of hybridized filters were analyzed using Phoretix Array professional software (Nonlinear Dynamics, Newcastle, UK). Images were imported as Tiff files and quantified by applying a grid composed of three columns and two rows of sub-grids on to the filter image, each grid containing 14 columns and 14 rows of spots. Spot alignment and spot editing manipulations were used to exclude artifacts from the analysis, and the intensity of each spot was quantified. An image rectangle was used for background subtraction. This method calculates the average intensity within a grid area where no specific hybridization occurred and then applies this as the background for each spot. The mean plus 2 SDs of the background measurements defined the detection level for the other spots on the array. Data normalization was performed by calculating the total signal detected within the gridded array and normalizing to the average volume. The normalization value was set to 100, enabling the calculated normalized values to appear as percentages of the total signal. Data were exported to Microsoft Excel and ratios between values for NM-2C5 and M-4A4 were calculated. Differential expression was highlighted if a ratio of more than threefold was apparent in either direction.

Quantitative PCR Analysis

Candidate gene mRNA transcripts were quantified in cultured cells and in primary xenograft tumor tissues recovered from athymic mice¹ using the ABI Prism 7700 Sequence Detection System (PE Applied Biosystems, Foster City, CA). Total RNA was isolated from homogenized cells/tissues using the RNeasy kit (Qiagen) and DNA was removed by digestion with RNase-free DNase-A treatment (Promega). cDNA was synthesized using MMLV reverse transcriptase and an oligo dT primer (Ambion, Austin, TX). PCR was performed using the SYBR Green PCR Master Mix kit containing SYBR Green I dye, AmpliTaq Gold DNA Polymerase, dNTPs with dUTP, passive reference and optimized buffer components (PE Applied Biosystems). PCR primers were designed against the 3'UTR of the human target genes using MacVector software (Oxford Molecular, Beaverton, OR) and designed to avoid potential binding to mouse homologue sequences. All primers were used at a final concentration of 100 nmol/L and 1 μ l of a 20- μ l cDNA reaction was added in 25- μ l PCR reactions. No-template controls were included for each target. Thermocycling was initiated with a 10-minute, 95°C enzyme activation step followed by 40 cycles of 95°C for 20 seconds, 60°C for 1 minute, and 72°C for 45 seconds. All reactions were performed in triplicate, and each reaction was verified to

contain a single product of the correct size using agarose gel electrophoresis. Data analysis was performed using the relative standard curve method as outlined by the manufacturer (PE Applied Biosystems) and as described previously.^{10,11} The mean GAPDH concentration (primer set supplied by PE Applied Biosystems) was determined once for each cDNA sample and used to normalize expression of all other genes tested in the same sample. The relative difference in expression was recorded as the ratio of normalized target concentrations for the same cDNA dilution. Sequence of selected primers: keratin 9 (accession no. NM.000226); forward primer 5'-CTTCTTCCTCAAATCTGGTGACC, reverse primer 5'-TCCAAAGC-CAAGCAGGGTTG; CD70 (accession no. NM.001252); forward 5'-CACTTTTGCCTTCCCGAAACAC, reverse 5'-CAATGCCTTCTTGTCTGCGCC; and TIMP3 (accession no. NM.000362) forward 5'-TTCTTCCCCACCTCAC-CATCTC reverse 5'-CTTCCTTCCCTCCCTCACTCTTAC. Sequence information of other PCR primers is available on request.

Results

cDNA Probe Hybridization and Analysis

To identify gene expressions that correlate with metastatic phenotype we performed comparative analysis of NM-2C5 and M-4A4 cells. Radioactively labeled first-strand cDNA probes synthesized from source mRNA were hybridized to Clontech Atlas Human Cancer cDNA arrays (Figure 1). Two concentrations of input RNA (100 ng and 1 μ g) were evaluated. Subsequent comparison of hybridization data with quantitative PCR of specific gene transcript abundance revealed that analysis of phosphorimages of hybridizations performed with probes derived from 1 μ g of polyA⁺ RNA were found to be optimal (Table 1). Image analysis revealed that the cDNA probes detected 139 (24%) of the 588 targets on the filter. As expected given the common origin of the two clones that comprise the metastasis model, the overall expression profiles were identical in pattern. Images from the two blots were processed using Phoretix software. Grids were created that matched the format of the duplicate spots on the array and normalized volume reports were exported to an Excel spreadsheet for further analysis. Only three genes were identified as being differentially expressed in this analysis. The tissue inhibitor of metalloproteinase-3 (TIMP3) and matrix metalloproteinase-8 genes were both estimated to be sixfold overexpressed in nonmetastatic NM-2C5 cells relative to M-4A4. Transcripts of the retinoic acid receptor- β family were estimated to be threefold more abundant in metastatic M-4A4 cells (Table 1).

Arbitrary Primer Design

There are few constraints on the design of arbitrary primers for RAP-array analysis. The primers need to have 40 to 50% C or G bases and to be designed to avoid complementary binding to themselves or the other primer

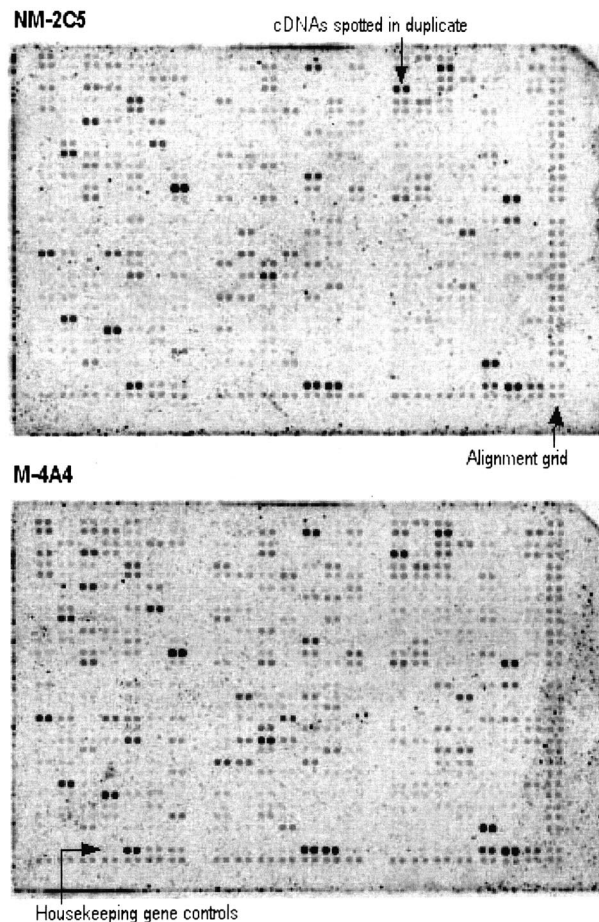


Figure 1. Imaging of Atlas Human Cancer cDNA arrays probed with ³³P-labeled cDNA from NM-2C5 and M-4A4 cells. Each array contains 588 cDNA fragments (10 ng of DNA per spot) spotted in duplicate. Negative controls and housekeeping genes are included, and genomic DNA spots serve as orientation markers. Actual size of arrays is 80 × 120 mm. mRNA from cultured NM-2C5 and M-4A4 cells was used to generate radiolabeled cDNA probes. Arrays were exposed to a phosphorimaging system for 2 days. Images from the two blots were processed using Phoretix Array software. Quantitative data from volume reports were exported for further processing. Analysis identified three differentially expressed genes described in Table 1.

in a pair. When using multiple fingerprints, the primers should also be sufficiently different to limit redundancy of transcript amplification. However, following these guidelines is not sufficient, the utility of the RAP-PCR primer combinations needs to be tested empirically. Because of the ability of the primers to bind in either direction some primers will produce reasonable fingerprints when used alone in the PCR reaction (Figure 2), but increased complexity and yield will result if two primers are used. Those primer combinations that produce a good yield and a wide range of product size are optimal (Figure 2). The visualization of the fingerprints in Figure 2 highlights the difficulty of identifying differentially expressed products using gel-based differential display methods alone. To check the fidelity of this method with our cell line mRNA samples we designed one primer pair (pair C) to bind to transcripts of the CTGF gene that we knew to be expressed approximately sevenfold higher in NM2C5 cells relative to M-4A4.

Table 1. Genes Identified as Differentially Expressed between NM-2C5 and M-4A4 Cells

Differentially expressed genes	NCBI reference	Fold change	
		cDNA probe	q-PCR
Up-regulated in NM-2C5			
Tissue inhibitor of metalloproteinase 3 (TIMP3)	NM_000362	6.34	7.0
Matrix metalloproteinase 8 (MMP8)	NM_002424	6.0	>20
Up-regulated in M-4A4			
Retinoic acid receptor β (RXR- β)	NM_021976	3.23	2.2

Radiolabeled cDNA probes were hybridized independently to Atlas Human Cancer cDNA arrays (Figure 1). Genes found to be up-regulated in each cell line relative to the other are listed. Fold-change levels refer to the magnitude of differential expression according to normalized array signals. Differential expression levels were verified in independent mRNA source material using real-time quantitative PCR (qPCR).

RAP-Array Analysis

Our primary goal was to use the RAP-array technique to identify differentially expressed genes, but we also used the established breast metastasis model to evaluate and optimize the utility of the analytical approach. Hybridization yields and patterns achieved with five single-fingerprint RAP probes (Figure 3) were compared with those achieved with complex cDNA probes. Hybridization of single fingerprint probes ranged from 61 (10.4%) to 116 (19.7%) of the 588 filter array targets. The sum of five independent hybridizations using distinct primer pairs resulted in up to 414 genes being detected (a 70% coverage). However, being arbitrarily designed, distinct primer pairs can bind to and amplify the same targets in a given cDNA population. Pairwise comparisons of the redundancy between single RAP-PCR fingerprints revealed target overlaps of up to 20%. A total of 344 (57%) target genes were detected by five single RAP-PCR fingerprints in individual hybridizations. Thirty-eight (27%) of the 139 genes detected by the cDNA probe were detected by the RAP-PCR probes. Therefore, sequential application of RAP-PCR probes allowed the screening of a greater number and of an alternative fraction of the transcript population than was achieved with a radiolabeled total cDNA probe. Analysis of the RAP-array experiments revealed a number of differentially expressed

genes. Using an arbitrary cutoff threshold of a threefold change between the two samples, nine genes were identified as being differentially expressed (Table 2). Of these, *CTGF* was expected to be detected as the primers used in fingerprint C were specifically designed to amplify the *CTGF* gene (indicated in Figure 4) that we knew to be differentially expressed through previous, unrelated analyses (data not shown). Having been identified as one of three genes differentially expressed in the cDNA probe experiments, the *TIMP3* gene was also identified by RAP-array analysis.

One way to increase array coverage per hybridization and, therefore, increase sample throughput in RAP-array analyses would be to combine labeled probes before hybridization. However, the advantage of reduced probe complexity could be lost with this strategy. To evaluate this we investigated whether the combination of up to five RAP-probes would compromise the ability of the RAP-array approach to identify differentially expressed genes.

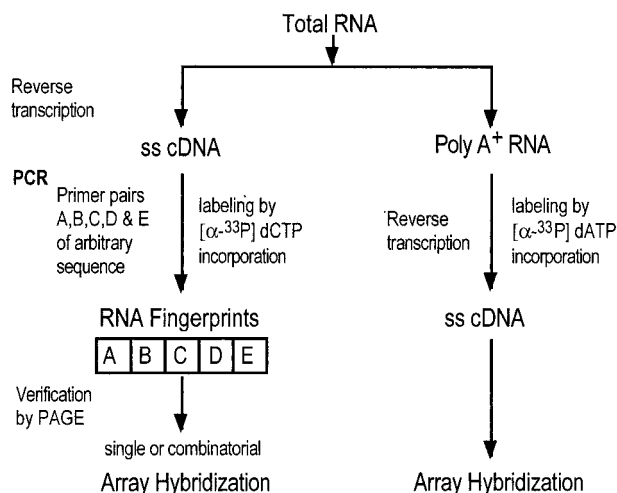


Figure 2. Flow chart depicting the steps involved in cDNA and RAP-PCR probe production. ss, Single-stranded; PAGE, polyacrylamide gel electrophoresis.

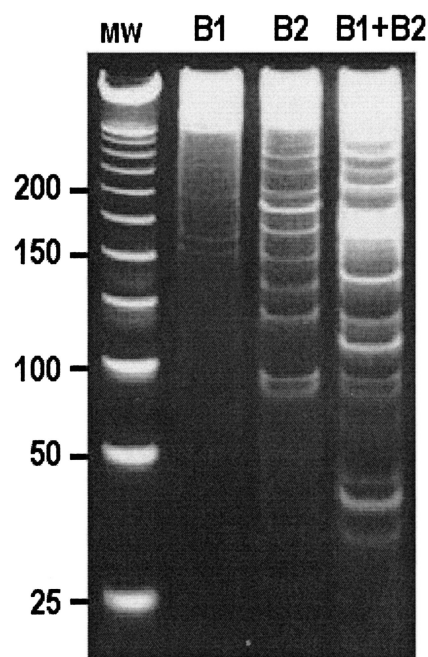


Figure 3. RAP-PCR fingerprints resolved by acrylamide gel electrophoresis. mRNA from NM-2C5 cells was reverse-transcribed and subjected to PCR using arbitrary primers B1 and B2. Single primers amplified cDNA with differing efficiencies (B1 and B2). The presence of both primers in a reaction resulted in a fingerprint of greater yield and complexity (B1+B2). MW, Molecular weight markers.

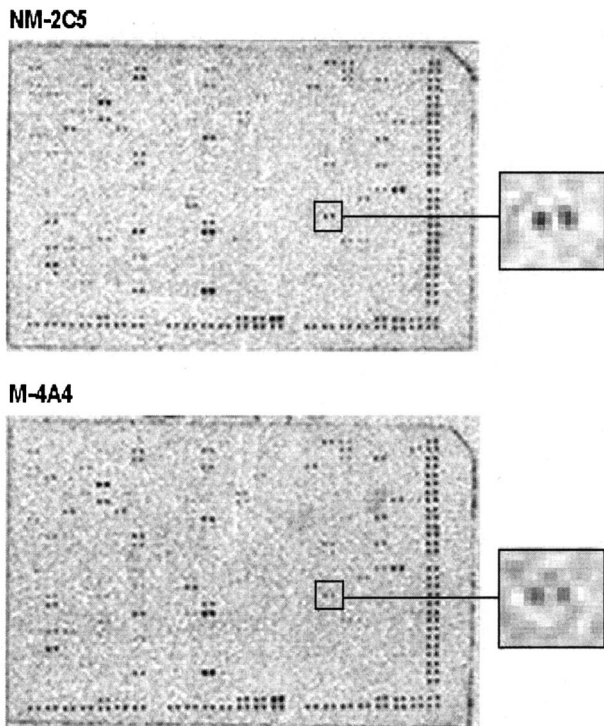


Figure 4. Imaging of Atlas Human Cancer cDNA arrays probed with a ³²P-labeled RAP-PCR fingerprint. mRNA from cultured NM-2C5 (**top**) and M-4A4 cells was used to generate a radiolabeled probe, in this case Fingerprint-C. Arrays were exposed to a phosphorimaging system for 2 days. An example of a differentially expressed target gene (*CTGF*) is **boxed** and expanded (**inset**). Images from the two blots were processed using Phoretix Array software. Grid objects were created and used for simultaneous quantitation of all duplicate spots on each array. Background subtracted and normalized volume reports were exported for further processing. Differentially expressed genes detected are described in Table 2.

We found that an interference factor is introduced when combining various pairs of labeled RAP-PCR probes. RAP-PCR fingerprints A and B highlighted 116 (19.7%) and 93 (15.8%) genes on the array, respectively, when used singly. Because 38 of the total 209 genes revealed by these two fingerprints were redundant, it would be

expected that a total of 171 genes would be revealed when the probes were combined before hybridization. This particular combination hybridized to only 115 genes, a reduction of ~33%. A similar trend was observed with other RAP-PCR probe combinations and the interference progressively increased as more probes were combined in a single hybridization. The combination of all five probes in one hybridization revealed 208 genes on the array. Allowing for redundancy between single probe hybridizations, 344 independent target genes were revealed when the five probes were used singly. Thus, a reduction of ~40% target detection occurred on combination of the five RAP-PCR probes. Furthermore, some of the specific differential expression revealed using single probe hybridizations was lost on probe combination. For example, expression of *CDK2* was revealed to be fivefold higher in M-4A4 cells by single RAP-PCR probe D, but this was reduced to a 2.3-fold difference (below the cutoff threshold) when analyzed using the five probes combined. The combination of RAP-PCR probes before array hybridization diminishes the target coverage and reduces the discriminatory power with regard to differential expression between samples.

Verification of Candidate Gene Expression in Vitro and in Vivo

One advantage of RAP-array over conventional differential display techniques is the immediate identification of differentially expressed genes, and thus sequence information of candidates is available without cloning procedures. In this study, real-time PCR (qPCR) was performed to verify and accurately quantitate the relative levels of each candidate gene. For each gene, specific primers were designed to operate under more stringent conditions than that of the RAP-PCR fingerprint synthesis. For this reason, the qPCR data are more accurate than the estimates achieved through image analysis. Of the nine candidate genes revealed by RAP-array analysis to have

Table 2. Differentially Expressed Genes of NM-2C5 and M-4A4 Cells Using RAP-Array

Differentially expressed genes	NCBI reference	Fold change				
		RAP-PCR fingerprints (FP)			q-PCR	
		Single FP (A to E)	A+B	A+B+C+D+E	<i>In vitro</i>	<i>In vivo</i>
Up-regulated in NM-2C5						
CCND1: cyclin D1	NM_053056	3.12 (E)	n/a	0.99	1.0	nd
CTGF: connective tissue growth factor	NM_001901	3.26 (C)	n/a	3.65	7.0	10.0
CCNA2: cyclin A2	NM_001237	4.11 (B)	1.67	1.45	2.0	nd
KRT9: keratin 9	NM_000226	17.67 (B)	7.48	5.98	>100	5.0
TIMP3: tissue inhibitor of metalloproteinase 3	NM_000362	4.72 (B)	3.8	3.4	7.0	1.0
MT3: metallothionein 3	NM_005954	3.82 (A)	0.58	1.09	2.8	1.5
Up-regulated in M-4A4						
CDK2: cyclin-dependent kinase 2	NM_001798	5.03 (D)	n/a	2.26	4.0	1.2
ERBB4: HER4	NM_005235	4.1 (C)	n/a	1.02	1.0	nd
CD70: CD27-ligand	NM_001252	6.03 (E)	n/a	2.5	5.0	10.0

Genes found to be up-regulated in each cell line relative to the other on analysis of radiolabeled RAP-PCR fingerprint probe data (Figure 1) are listed. Single RAP-PCR fingerprints (FP) A through E were hybridized independently to Atlas Human Cancer cDNA arrays (A to E), and in some cases in combination (A+B) and (A+B+C+D+E). Fold-change levels refer to the magnitude of differential expression according to normalized array signals. Differential expression levels were verified in independent mRNA source material obtained from cells in culture (*in vitro*) and from three primary tumor xenografts (*in vivo*) using real-time quantitative PCR (qPCR). qPCR data are presented as the mean of triplicate reactions.

at least threefold differential expression between NM-2C5 and M-4A4 cells in either direction, six were confirmed by qPCR (Table 2). Among the genes identified in this study, matrix metalloproteinase-8² and CTGF were previously known to be expressed at higher levels in NM-2C5 over M-4A4 both *in vitro* and *in vivo*, but the finding that the CDK2, TIMP3, keratin 9, and CD70 genes were differentially expressed in these cells was novel (Table 2). We also investigated whether the relative levels of the most markedly differentially expressed genes were maintained in NM-2C5 and M-4A4 cells growing *in vivo* as a primary tumor xenograft.¹ Three tumors were recovered from orthotopically inoculated athymic mice for subsequent RNA extraction and PCR analysis. PCR primers were designed to amplify sequences specific to human transcripts to avoid a possible contribution of murine transcripts to the analysis. In the majority of cases this was achieved by using human-specific sequences located in the 3'-untranslated region of the target mRNA. Of the novel candidate genes tested, transcripts of the keratin 9 and CD-27 ligand genes were found to maintain their differential expression profile *in vivo* (Table 2).

Discussion

The development of DNA-array systems has greatly facilitated the high-throughput detection and comparative analysis of complex gene expression profiles. In this study we used DNA macroarrays created on nylon membranes containing 588 unique cDNAs formatted at relatively low density. These and other expression arrays are usually probed with labeled first-strand cDNA synthesized from a mRNA source of interest. However, this approach results in the production of a very complex probe that in this context is biased toward the detection of high-abundance transcripts.³ In RAP-PCR sampling, the prevalence of an amplified product is dependent on both the abundance of a transcript and the complementarity of the arbitrary primers with a transcript. This allows rare messages to be sampled among more abundant targets. This is an important advantage of RAP-PCR over other sampling methods.^{6,12}

In this study we used a RAP-array technique to perform comparative gene expression analyses between breast tumor cells that have opposite metastatic phenotypes.¹ To test the ability of the system to detect genes that were differentially expressed we compared data obtained using labeled cDNA probes and single- and multiple-RAP-PCR probes. The sequential application of five RAP-PCR fingerprints gave 57% coverage of the 588 targets on the Atlas Human Cancer cDNA array, a considerable improvement over the 24% gene coverage achieved with a direct labeled cDNA probe. Allowing for redundancy between RAP-PCR probes, 344 individual genes on the array were identified as being expressed in the breast tumor cell lines and nine genes were identified as being differentially expressed (fold-change more than threefold) between the isogenic metastatic and nonmetastatic counterparts. Of these nine genes, six were confirmed to be differentially expressed in cultured cells and two were

confirmed to maintain differential expression in primary tumor xenograft tissue. Considering the relatively simple format of the arrays used in this study, to identify two genes that we are confident are significantly differentially expressed in our model is a reasonable return. Even though expanding the experiment by creating more fingerprints leads to diminishing returns because of increasing target detection overlap, these data suggest that as few as 10 RAP-PCR fingerprints would sample the majority of transcripts in a given cell mRNA population.

Comparison of the RAP-array data with alternative array analyses we have performed previously (manuscript in preparation) reveals that the detection and verification rates are similar between approaches. Of the ~12,000 genes on a Affymetrix U95 chip, ~45% were deemed to be present in the transcriptome of NM-2C5 and M4A4 cells. Although only 24% of the 588 cDNAs on the Atlas array were revealed using a labeled cDNA probe, a 57% target detection rate was achieved by sequential hybridization of five RAP-PCR probes, a considerably higher coverage rate than that revealed using the GeneChip approach. These differences are most likely because of the amplification procedures used in each experiment. The biochemistry involved in Affymetrix GeneChip analysis includes a linear cRNA amplification that increases the probe yield for hybridization and subsequent detection, but still favors the detection of the more abundant transcripts within the complex probe. RAP-PCR sampling is not biased toward the abundant transcripts and thus can bring relatively rare messages into the detection range. However, in practice it is worth noting that some gene transcripts were detected in multiple RAP-PCR fingerprints, indeed the mRNA for vimentin was amplified in all five fingerprints. Although it is possible that this occurs by chance, the targets in question (vimentin, β -actin, and ribosomal protein S9) were among the most abundant transcripts within our samples and so it seems that there may be an amplification bias toward very abundant genes in RAP-PCR. Such amplification bias in an exponential reaction results in saturation of those targets on the array and so analysis of any potential differential expression of those targets will not be possible. In this study, RAP-PCR probes sampled both a greater and an alternative transcript population than that sampled by direct labeling of cDNA. The data suggest that the sequential use of a cDNA probe (for abundant genes) and five RAP-PCR probes is an effective approach, resulting in a combined 70% array target coverage.

The combination of RAP-PCR probes before hybridization would clearly be advantageous from the point of view of efficiency of coverage and, therefore, throughput. This study revealed that an experimental trade-off occurs with such probe combination. Combination of RAP-PCR probes did expectedly improve the target coverage achieved in a single hybridization, but a reduction of ~40% overall target detection occurred on combination of the five RAP-PCR probes. More importantly, much of the information regarding differential expression was lost when probes were combined. These data suggest that interference factors are introduced when probes are combined. The reduction in target coverage could be

because of interprobe hybridization in solution, and the loss of differential expression information may be because of the saturation of specific targets through hybridization of radiolabeled complementary sequences emanating from more than one RAP-PCR probe. It was concluded that to maximize the advantage of reduced probe complexity it is optimal to use the RAP-PCR probes in independent hybridizations.

Array-based gene expression profiling techniques do not necessarily give the investigator a definitive answer. Because of the multiple steps of sample manipulation involved in the production of labeled probes, the results are only estimates of actual transcript levels. Thus, array-based differential expression of specific genes requires verification. In this study, six (66%) of the candidate genes identified as being differentially expressed by array analysis alone were verified using qPCR. Although false-positive signals may occur because of random PCR artifacts or hybridization anomalies, the most common source of error in array-based analyses is the array itself. At a given location, there can be considerable interarray variation when more or less target DNA is applied to one array over another during manufacture.

Given the isogenic nature of the metastasis model cell lines,¹ the identification of any genes that are differentially expressed both *in vitro* and *in vivo* is significant. Our previous comparative studies^{1,2} have identified only two genes to be consistently overexpressed in metastatic M-4A4 cells and seven genes overexpressed in NM-2C5 cells grown *in vitro* and *in vivo*. The presented array analyses revealed four genes in this category, two that have been revealed in previous work (matrix metalloproteinase-8 and CTGF) and two that were previously unidentified. The CTGF gene acted as a valuable test target in these studies. By designing one of the PCR primer pairs to specifically amplify CTGF we were able to stringently test the utility of the RAP-array assay using previous knowledge. The RAP-array assay estimated CTGF expression as more than threefold higher in NM-2C5 cells relative to M-4A4, a finding reasonably consistent with the actual quantitative difference of approximately sevenfold between the cells grown in culture. This data indicates that RAP-PCR coupled with detection using the Atlas Human Cancer cDNA expression array can reliably identify differentially expressed genes. It is notable that on the same array the cDNA probe did not detect CTGF expression in either cell line. The novel finding was the differential expression of the keratin 9 and CD70 ligand genes. The expression of the keratin 9 gene was particularly impressive. The fold-change revealed by RAP-array was >17-fold, and verification by qPCR revealed that this was a considerable underestimate of the differential expression in cultured cells. Furthermore, both genes were confirmed to be differentially expressed by at least fivefold in RNA samples recovered from the cells growing *in vivo*, making them worthy of further investigation with regard to a possible role in the metastatic phenotype. It is noteworthy that neither of these differences were identified using the cDNA probe, nor in previously used alternative comparative screening techniques. Both *keratin 9* and *CD70* sequences are present as targets on the Affymetrix U95

GeneChip, but while confirmed as being expressed in both cell lines using that format, neither were revealed as differentially expressed.

The keratin gene family encodes intermediate filaments that are a major component of the mammalian cytoskeleton, and are mainly expressed in cells of epithelial origin.¹³ The type I keratin 9 is reported to be expressed only in the terminally differentiated epidermis of the palms and soles. Mutations in this gene cause epidermolytic palmoplantar keratoderma.¹⁴ To our knowledge keratin 9 has not specifically been linked with human cancer previously, but the co-expression of keratin-type intermediate filaments and vimentin in the same cell has been reported to correlate with increased invasiveness and a more aggressive tumorigenic phenotype.^{15,16} Furthermore, in the presence of vimentin, specific keratins act to stabilize others within the same gene family.¹⁷ Vimentin is an intermediate filament protein found predominantly in cells of parenchymal origin¹⁸ but many transformed cell lines have been shown to express vimentin regardless of phenotype.¹⁹ Indeed, both NM-2C5 and M-4A4 cells express large amounts of vimentin and thus it may be the balance of keratin-type gene expression in these cells that results in the induction or suppression of their motile and invasive phenotype. The protein encoded by the CD70 gene (also known as TNFSF7, CD27 ligand, and Ki-24 antigen) is a cytokine that belongs to the tumor necrosis factor (TNF) ligand family.^{20,21} In contrast to the expression of other TNF/TNFR family members, expression of CD70 and its receptor CD27 is predominantly confined to lymphocytes.²² Although CD70 expression has not been well documented in nonlymphoproliferative disorders, Wischhusen and colleagues²³ have postulated that induction of B-cell and T-cell apoptosis via interactions of CD70 expressed on glioma cells and CD27 expressed on B and T cells may be a novel way for the immune escape of malignant gliomas. It is conceivable that CD70 plays a role in regulating B-cell activation and cytotoxic function of natural killer cells in the athymic mice used to assess the metastatic propensity of M-4A4 cells. Manipulation of the keratin 9 and CD70 candidate genes in the respective counterpart of our metastasis model and subsequent analysis of upstream and downstream effects will permit evaluation of their functional significance in the metastatic phenotype and will elucidate the intracellular molecular networks in which these genes participate.

In summary, there are several advantages of the RAP-array technique over other comparative analyses. The inclusion of a PCR amplification means that it is possible to use samples of as little as 500 ng of total RNA. The hybridization of the RAP-PCR fingerprints to arrays of known DNA targets avoids cloning and sequencing regimes and enhances detection by separating hybridization events for image processing. There are many modifications that one can envisage for further improvement of the RAP-array technique. Greater target coverage can be attained by using arrays with more targets, and perhaps even customized arrays that contain targets complementary to transcripts known to be amplified by a set of arbitrary primers. Techniques evolving within the cDNA

array field, such as labeling with fluorescent dyes for improved sensitivity and more linear detection, apply equally to RAP-array analyses. Within this study, a combination of cDNA and RAP-PCR probes gave the best coverage of targets and sampled transcripts across a broad range of abundance within a complex mixture. Therefore, although the method has its advantages, the RAP-array technique is perhaps best used as an adjunct to other, more high-throughput array screening methods.

Acknowledgments

We thank Drs. Mike McClelland, John Welsh, and Gaelle Rondeau of the Sidney Kimmel Cancer Center, San Diego, CA, for technical advice and discussion; and Doug Bodde of Non-Linear Dynamics for help and advice with image data analysis.

References

1. Urquidi V, Sloan D, Kawai K, Agarwal D, Woodman AC, Tarin D, Goodison S: Contrasting expression of thrombospondin-1 and osteopontin correlates with absence or presence of metastatic phenotype in an isogenic model of spontaneous human breast cancer metastasis. *Clin Cancer Res* 2002, 8:61–74
2. Agarwal D, Goodison S, Nicholson B, Tarin D, Urquidi V: Expression of matrix metalloproteinase 8 (MMP-8) and tyrosinase-related protein-1 (TYRP-1) correlates with the absence of metastasis in an isogenic human breast cancer model. *Differentiation* 2003, 71:114–125
3. Boll W, Fujisawa J, Niemi J, Weissmann C: A new approach to high sensitivity differential hybridization. *Gene* 1986, 50:41–53
4. Ramsay G: DNA chips: state-of-the art. *Nature Biotechnol* 1998, 16:40–44
5. Marshall A, Hodgson J: DNA chips: an array of possibilities. *Nature Biotechnol* 1998, 16:27–31
6. Trenkle T, Welsh J, Jung B, Mathieu-Daude F, McClelland M: Non-stoichiometric reduced complexity probes for cDNA arrays. *Nucleic Acids Res* 1998, 26:3883–3891
7. Welsh J, McClelland M: Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Res* 1990, 18:7213–7218
8. Welsh J, Chada K, Dalal SS, Cheng R, Ralph D, McClelland M: Arbitrarily primed PCR fingerprinting of RNA. *Nucleic Acids Res* 1992, 20:4965–4970
9. McClelland M, Ralph D, Cheng R, Welsh J: Interactions among regulators of RNA abundance characterized using RNA fingerprinting by arbitrarily primed PCR. *Nucleic Acids Res* 1994, 22:4419–4431
10. Schmittgen TD, Zakrajsek BA, Mills AG, Gorn V, Singer MJ, Reed MW: Quantitative reverse transcription-polymerase chain reaction to study mRNA decay: comparison of endpoint and real-time methods. *Anal Biochem* 2000, 285:194–204
11. Rajeevan MS, Vernon SD, Taysavang N, Unger ER: Validation of array-based gene expression profiles by real-time (kinetic) RT-PCR. *J Mol Diagn* 2001, 3:26–31
12. Trenkle T, Mathieu-Daude F, Welsh J, McClelland M: Reduced complexity probes for DNA arrays. *Methods Enzymol* 1999, 303:380–392
13. Porter RM, Lane EB: Phenotypes, genotypes and their contribution to understanding keratin function. *Trends Genet* 2003, 19:278–285
14. Kobayashi S, Tanaka T, Matsuyoshi N, Imamura S: Keratin 9 point mutation in the pedigree of epidermolytic hereditary palmoplantar keratoderma perturbs keratin intermediate filament network formation. *FEBS Lett* 1996, 386:149–155
15. Kirschmann DA, Seftor EA, Nieva DR, Mariano EA, Hendrix MJ: Differentially expressed genes associated with the metastatic phenotype in breast cancer. *Breast Cancer Res Treat* 1999, 55:127–136
16. Chu YW, Seftor EA, Romer LH, Hendrix MJ: Experimental coexpression of vimentin and keratin intermediate filaments in human melanoma cells augments motility. *Am J Pathol* 1996, 148:63–69
17. Pankov R, Simcha I, Zoller M, Oshima RG, Ben-Ze'ev A: Contrasting effects of K8 and K18 on stabilizing K19 expression, cell motility and tumorigenicity in the BSp73 adenocarcinoma. *J Cell Sci* 1997, 110:965–974
18. Steinert PM, Liem RK: Intermediate filament dynamics. *Cell* 1990, 60:521–523
19. Helftenbein G, Alvarez CV, Tuohimaa P, Beato M: Expression of epithelial phenotype is enhanced by v-Ha-ras in rat endometrial cells immortalized by SV40 T antigen. *Oncogene* 1993, 8:2075–2085
20. Goodwin RG, Alderson MR, Smith CA, Armitage RJ, VandenBos T, Jerzy R, Tough TW, Schoenborn MA, Davis-Smith T, Hennen K, et al: Molecular and biological characterization of a ligand for CD27 defines a new family of cytokines with homology to tumor necrosis factor. *Cell* 1993, 73:447–456
21. Bowman MR, Crimmins MA, Yetz-Aldape J, Kriz R, Kelleher K, Herrmann S: The cloning of CD70 and its identification as the ligand for CD27. *J Immunol* 1994, 152:1756–1761
22. Lens SM, Baars PA, Hooibrink B, van Oers MH, van Lier RA: Antigen-presenting cell-derived signals determine expression levels of CD70 on primed T cells. *Immunology* 1997, 90:38–45
23. Wischhusen J, Jung G, Radovanovic I, Beier C, Steinbach JP, Rimner A, Huang H, Schulz JB, Ohgaki H, Aguzzi A, Rammensee HG, Weller M: Identification of CD70-mediated apoptosis of immune effector cells as a novel immune escape pathway of human glioblastoma. *Cancer Res* 2002, 62:2592–2599