Nucleic Acids Research, 1998, Vol. 26, No. 17 **3883–3891**

Non-stoichiometric reduced complexity probes for cDNA arrays

Thomas Trenkle, John Welsh*, Barbara Jung, Francoise Mathieu-Daude and Michael McClelland

Sidney Kimmel Cancer Center, 10835 Altman Row, San Diego, CA 92121, USA

Received June 23, 1998; Revised and Accepted July 9, 1998

DDBJ/EMBL/GenBank accession no. AF067817

ABSTRACT

A method is presented in which the reduced complexity and non-stoichiometric amplification intrinsic to RNA arbitrarily primed PCR fingerprinting (RAP-PCR) is used to advantage to generate probes for differential screening of cDNA arrays. RAP-PCR fingerprints were converted to probes for human cDNA clones arrayed as *Escherichia coli* colonies on nylon membranes. Each array contained 18 432 cDNA clones from the IMAGE consortium. Hybridization to ~1000 cDNA clones was detected using each RAP-PCR probe. Different RAP-PCR fingerprints gave hybridization patterns having very little overlap (<3%) with each other or with hybridization patterns from total cDNA probes. Consequently, repeated application of 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. This method was applied to RNA from HaCaT keratinocytes treated with epidermal growth factor. Two RAP-PCR probes detected hybridization to 2000 clones, from which 22 candidate differentially expressed genes were observed. Differential expression was tested for 15 of these clones using RT–PCR and 13 were confirmed. The use of this cDNA array to analyze RAP-PCR fingerprints allowed for an increase in detection of 10–20-fold over the conventional denaturing polyacrylamide gel approach to RAP-PCR or differential display. Throughput is vastly improved by the reduction in cloning and sequencing afforded by the use of arrays. Also, repeated cloning and sequencing of the same gene or of genes already known to be regulated in the system of interest is minimized. The procedure we describe uses inexpensive arrays of plasmid clones spotted as *E.coli* colonies to detect differential expression, but these reduced complexity probes should also prove useful on arrays of PCR-amplified fragments and on oligonucleotide chips. Genes observed in this manuscript: H11520, U35048, R48633, H28735, M13918, H12999, H05639, X79781, M31627, H23972, AB000712, R75916, U66894, AF067817.

INTRODUCTION

Arrays of cDNA clones or oligonucleotides affixed to a solid support can capture labeled homologous cDNA from solution and, thereby, measure the differential expression of many genes in parallel. However, a total labeled cDNA probe from a mammalian cell typically has a complexity of >30 000 000 bases, which complicates attempts to detect differential expression among the rarer mRNAs using differential hybridization. Recent advances in the use of fluorescence and confocal microscopy have led to improvements in the sensitivity and dynamic range of differential hybridization methods and the detection of transcripts at a sensitivity approaching 1/500 000 (1,2 and references therein).

Despite these improvements, several of these methods are currently too expensive for the average molecular biology laboratory to implement. On the other hand, arrays of *Escherichia coli* colonies containing tens of thousands of sequenced ESTs are available for differential screening and are quite inexpensive. The standard method for differential screening, which typically uses probes derived from reverse transcription of total message and autoradiography or phosphorimaging, can give impressive results (3). However, the method is limited to the most abundant messages; only these abundant transcripts are represented highly enough to yield effective probes with a sensitivity of perhaps 1/15 000 (4). Here we show that differential screening of arrays of plasmids in colonies can be improved greatly by reducing the complexity of the probe and by systematically increasing the contribution of rarer mRNAs to the probe. In this way, differential screening using these arrays is not confined to only the most abundant mRNAs.

One way to construct a probe having reduced complexity and increased representation of rare messages is to use RAP-PCR fingerprinting, which samples a reproducible subset of the message population based on the best matches with arbitrary primers (5,6). In a typical RAP-PCR fingerprint, ~50–100 cDNA fragments per lane are visible on a polyacrylamide gel, including products from relatively rare mRNAs that happen to have among the best matches with the arbitrary primers. If only 100 cDNA clones could be detected in an array by each probe, then hybridization to arrays would be inefficient. However, RAP-PCR fingerprints contain many products that are too rare to visualize by autoradiography of a polyacrylamide gel. Nonetheless, these

*To whom correspondence should be addressed. Tel: +1 619 450 5990; Fax: +1 619 550 3998; Email: jwelsh@skcc.org

rarer products are reproducible and of sufficient abundance to serve as probe for arrays when labeled at high specific activity.

The experiments presented here show that a single probe derived from RAP-PCR can detect ~1000 cDNAs on an array containing ~18 432 EST clones, a 10–20-fold improvement over the performance of fingerprints displayed on denaturing polyacrylamide gels. In addition, when a differentially regulated gene is detected on a cDNA array, a clone representing the transcript is immediately available and often sequence information for the clone is also available. Furthermore, the clones are usually much longer than the usual RAP-PCR product. In contrast, the standard approaches to RNA fingerprinting require that the product be gel purified and sequenced before verification of differential expression can be performed.

In this report, we show that expression differences that can be seen in a standard RAP-PCR fingerprint can also be detected using fingerprints as differential screening probes against arrays. We further show that differentially amplified RAP-PCR products that are below the detection capabilities of the standard denaturing polyacrylamide gel and autoradiography methods can be detected using hybridization to cDNA arrays.

MATERIALS AND METHODS

RNA preparation

The immortal human keratinocyte cell line HaCaT (7) was grown to confluence and maintained at confluence for 2 days. The medium (Dulbecco's modified Eagle's medium + 10% fetal bovine serum + penicillin/streptomycin) was changed 1 day prior to experiments. Epidermal growth factor (EGF) (Gibco-BRL) was added at 20 ng/ml or transforming growth factor TGF-β (R&D systems, Minneapolis, MN) was added at 5 ng/ml. Treated and untreated cells were harvested after 4 h by scraping the Petri dishes in the presence of lysis buffer (Qiagen, Chatsworth, CA) and homogenized through Qiashredder columns. On average, 7×10^6 cells (confluent growth in a 100 mm diameter Petri dish) yielded 40 µg total RNA from the RNeasy total RNA purification kit (Qiagen). RNA, in 20 mM Tris, 10 mM $MgCl_2$ buffer, was incubated with 0.08 U/µl RNase-free DNase and 0.32 U/µl RNase inhibitor (both from Boehringer Mannheim Biochemicals, Indianapolis, IN) for 40 min at 37°C and cleaned again using the RNeasy kit. This step is important because small amounts of genomic DNA can contribute to the fingerprints. RNA quantity was measured by spectrophotometry and RNA samples were adjusted to 400 ng/µl in water. They were checked for quality and concentration by agarose gel electrophoresis and stored at –20°C.

RNA fingerprinting

RAP-PCR was performed using standard protocols (8,9). Reverse transcription was performed on total RNA using four concentrations per sample (1000, 500, 250 and 125 ng/ reaction) and an oligo(dT) primer (15mer) (Genosys Biotechnologies, The Woodlands, TX). RNA (5 µl) was mixed with 5 µl RT mixture for a 10 µl final reaction containing 50 mM Tris, pH 8.3, 75 mM KCl, 3 mM $MgCl_2$, 20 mM DTT, 0.2 mM each dNTP, 0.5 µM primer and 20 U MuLV reverse transcriptase (Promega, Madison, WI). RNA samples were checked for DNA contaminants by including a reverse transcriptase-free control in initial RAP-PCR experiments. The reaction was performed at 37°C for 1 h (after a 5 min ramp from 25 to 37°C), the enzyme was inactivated by heating the samples at 94°C for 5 min and the newly synthesized cDNA was diluted 4-fold in water.

PCR was performed after the addition of a pair of two different 10mer or 11mer oligonucleotide primers of arbitrary sequence: pair A, GP14 (GTAGCCCAGC) and GP16 (GCCACCCAGA); pair B, Nuc1+ (ACGAAGAAGAAGAG) and OPN24 (AGGGGCACCA). In general, there are no particular constraints on the primers except that they contain at least a few C or G bases, that the 3′-ends are not complementary with themselves or the other primer in the reaction, to avoid primer dimers, and that primer sets are chosen that are different in sequence so that the same parts of mRNA are not amplified in different fingerprints.

Diluted cDNAs (10 µl) were mixed with the same volume of 2× PCR mixture containing 20 mM Tris, pH 8.3, 20 mM KCl, 6.25 mM $MgCl_2$, 0.35 mM each dNTP, 2 µM each oligonucleotide primer, 2 µCi [α-^{32}P]dCTP (ICN, Irvine, CA) and 5 U Ampli*Taq*® DNA polymerase Stoffel fragment (Perkin-Elmer Cetus, Norwalk, CT) for a 20 µl final reaction. Thermocycling was performed using 35 cycles of 94°C for 1 min, 35°C for 1 min and 72°C for 2 min.

An aliquot of the amplification products (3.5 µl) was mixed with 9 µl formamide dye solution, denatured at 85°C for 4 min and chilled on ice. A sample of 2.4 µl was loaded onto a 5% polyacrylamide, 43% urea gel, prepared with 1× TBE buffer. The PCR products resulting from the four different concentrations of the same RNA template were loaded side by side on the gel (Fig. 1).

Electrophoresis was performed at 1700 V or at a constant power of 50–70 W until the xylene cyanol tracking dye reached the bottom of the gel (~4 h). The gel was dried under vacuum and placed on Kodak BioMax X-Ray film for 16–48 h.

Labeling of RAP-PCR products for use as probes against cDNA arrays

Up to 10 µg PCR product from RAP-PCR can be purified using a QIAquick PCR Purification Kit (Qiagen, Chatsworth, CA), which removes unincorporated bases, primers and primer dimers <40 bp. The DNA was recovered in 50 µl 10 mM Tris, pH 8.3.

Random primed synthesis with incorporation of [α-^{32}P]dCTP was performed using a standard protocol. Ten percent of the recovered fingerprint DNA (typically ~100 ng in 5 µl) was combined with 3 µg random hexamer oligonucleotide primer and 0.3 µg each of the fingerprint primers in a total volume of 14 µl, boiled for 3 min and then placed on ice.

The hexamer/primer/DNA mix was mixed with 11 µl reaction mix to yield a 25 µl reaction containing 0.05 mM three dNTPs (minus dCTP), 50 µCi [α-^{32}P]dCTP (3000 Ci/mmol, 5 µl), 1× Klenow fragment buffer (50 mM Tris–HCl, 10 mM $MgCl_2$, 50 mM NaCl, pH 8.0, and 4 U Klenow fragment; Gibco-BRL Life Technologies, Gaithersburg, MD). The reaction was performed at room temperature for 4 h. For maximum probe length the reaction was chased by adding 1 µl 1.25 mM dCTP and incubated for 15 min at 25°C, then for an additional 15 min at 37°C. The unincorporated nucleotides, hexamers and primers were removed with the Qiagen Nucleotide Removal Kit and the purified products were eluted using two aliquots of 140 µl 10 mM Tris, pH 8.3.

Labeling of poly(A)$^+$ mRNA and genomic DNA for use in arrays

Poly(A)$^+$-selected mRNA and genomic DNA were labeled using random hexamers. Genomic DNA (150 ng) was labeled using the

same protocol used for labeling the RAP-PCR products. Poly(A)$^+$ mRNA (1 µg) and 9 µg random hexamer in a volume of 27 µl were incubated at 70°C for 2 min and chilled on ice. The RNA/hexamer mix was mixed with 23 µl master mix containing 10 µl 5× AMV reaction buffer (250 mM Tris–HCl, pH 8.5, 40 mM $MgCl_2$, 150 mM KCl and 5 mM DTT), 1 µl 33 mM each of three dNTPs (minus dCTP), 2 µl AMV reverse transcriptase (Boehringer Mannheim, Indianapolis, IN) and 10 µl [α-^{32}P]dCTP (3000 Ci/mmol) in a final volume of 50 µl. The reaction was incubated at room temperature for 15 min, ramped for 1 h to 47°C, held at 47°C for 1 h and chased with 1 µl 33 mM dCTP for another 30 min at 47°C. The labeled products were purified as described above.

Hybridization to the array

When radioactivity is used to label the probe, four membranes are needed, one membrane for each of two concentrations of RNA for each of the two RNA samples to be compared. If two color fluorescence were to be used, then two arrays would be needed, one for each of the two concentrations of starting RNA, because probes from the two RNA samples can be mixed. Here we present a protocol for radiolabeled probes.

Prewash of cDNA filters. The cDNA filters (Genome Systems, St Louis, MO) were washed in three changes of 2× SSC, 0.1% SDS in a horizontally shaken flat bottomed container to reduce the residual bacterial debris. The first wash was carried out in 500 ml for 10 min at room temperature. The second and third washes were carried out in 1 l prewarmed (55°C) prewash solution for 10 min each.

Prehybridization. The filters were transferred to roller bottles and prehybridized in 60 ml prewarmed (42°C) prehybridization solution containing 6× SSC, 5× Denhardt's reagent, 0.5% SDS, 100 µg/ml fragmented denatured salmon sperm and 50% formamide for 1–2 h at 42°C in an oven.

Hybridization. The prehybridization solution was exchanged with 7 ml prewarmed (42°C) hybridization solution containing 6× SSC, 0.5% SDS, 100 µg/ml fragmented denatured salmon sperm and 50% formamide. To decrease the background hybridization due to repeats (e.g. Alu and Line elements), sheared human genomic DNA was denatured in a boiling water bath for 10 min and immediately added to the hybridization solution to a final concentration of 10 µg/ml. An aliquot of 10 ng/ml poly(dA) can be added to block oligo(dT) stretches in the radiolabeled probe. Simultaneously, the labeled probe, in a total volume of 280 µl, was denatured in a boiling water bath for 4 min and immediately added to the hybridization solution. The hybridization was carried out at 42°C for 2–48 h (typically 18 h) in large roller bottles.

Wash. For the washes the incubator oven temperature was set to 68°C. The hybridization solution was poured off and the membrane was washed twice with 50 ml 2× SSC, 0.1% SDS at room temperature (RT) for 5 min. The wash solution was then replaced with 100 ml 0.1× SSC, 0.1% SDS (RT) and incubated for 10 min. Further washes were performed in 100 ml 0.1× SSC, 0.1% SDS at 55–68°C for 40 min in the roller bottles, followed by washing in 1 l for 20 min with gentle agitation in a horizontal shaker. The filters were transferred back to the roller bottles containing 100 ml prewarmed (55–68°C) 0.1× SSC, 0.1% SDS and incubated for 1 h. The final wash solution was removed and the filters were briefly rinsed in 2× SSC at room temperature.

After washing, the membranes were blotted with 3MM paper, wrapped in Saran wrap while moist and placed against X-ray film. The membranes were usually sufficiently radioactive that a 1 day exposure with a screen will reveal the top 1000 products on an array of 18 432 bacterial colonies carrying EST clones. Weaker probes or fainter hybridization events can be seen using an intensifying screen at –70°C for a few days. Also, membranes may be read using a phosphorimager or using a fluorescence scanner when fluorescent probes are used.

Confirmation of differential expression using low stringency RT–PCR

The first level of confirmation is the use of two RNA concentrations per sample. Only those hybridization events that seem to indicate differential expression at both RNA concentrations in both RNA samples can be relied upon.

More than 70% of the IMAGE consortium clones have single pass sequence reads from the 5′- or 3′-end or both deposited in the GenBank database. In cases where there is no prior sequence information available, the clones can be ordered from Genome Systems and sequenced. Sequences were used to derive PCR primers of 18–25 bases in length using MacVector 6.0 (Oxford Molecular Group, Oxford, UK). Generally, primers were chosen to generate PCR products of 50–250 bp and have melting temperatures of at least 60°C.

Reverse transcription was performed under the same conditions as in the RAP-PCR protocol (above), using an oligo(dT) primer or a mixture of random 9mer primers (Genosys). The PCR reaction was performed using two specific primers (18–25mer). The PCR conditions were the same as in the RAP-PCR fingerprint protocol but 1.5 µM each primer was used. The following low stringency thermal profile was used: 94°C for 40 s, 47°C for 40 s and 72°C for 1 min, for 19, 22 and 25 cycles. The reactions were carried out in three sets of tubes at different cycle numbers because the abundance of the transcripts, the performance of the primer pairs and the amplifiability of the PCR products can vary. PCR products were run under the same conditions as above on a 5% polyacrylamide, 43% urea gel. The gel was dried and exposed for 18–72 h. Invariance among the other arbitrary products in the fingerprint was used as an internal control to indicate the reliability of the relative quantitation. Primer pairs (Genosys) used for confirmation of differential expression were as follows: GenBank accession no. H11520 (90 nt product), (A) AATGAGGGGGACAAATGGGAAGC, (B) GGAGAGCCCTTCCTCAGACATGAAG; TSC-22 gene (U35048, H11073, H11161) (179 nt product), (A) TGACAAAATGGTGACAGGTAGCTGG, (B) AAGTCCACACCTCCTCAGACAGCC; R48633 (178 nt product), (A) CCCAGACACCCAAACAGCCGTG, (B) TGGAGCAGCCGTGTGTGCTG. Figure 3 was assembled using Adobe PhotoShop.

RESULTS AND DISCUSSION

Choice of array

Arrays containing cDNA clones are available on nylon membranes from a variety of suppliers, including Research Genetics (www.resgen.com), Genome Systems (www.genomesystems.com) and the German Human Genome Project (www.rzpd.de). These arrays include clones from various human tissues, stages of

development and disease states. Arrays of mouse and yeast sequences are also available. At present, there are two types of arrays available on nylon membranes. One type contains 18 432 *E.coli* colonies, each carrying a different IMAGE consortium EST plasmid (www-bio.llnl.gov/bbrp/image/image.html), spotted twice on a 22 × 22 cm membrane (available from Genome Systems). The second type contains >5000 PCR products from selected IMAGE clones amplified using vector primers, available from Research Genetics. To date, an array of PCR products is available for every yeast ORF and for a subset of human ESTs. One can expect a dramatic increase in the number of available arrays, organisms and accompanying sequence information.

We chose the Genome Systems arrays, which contain by far the largest number of ESTs per unit cost. However, each spotted EST is associated with *E.coli* genomic DNA from the host, in contrast to PCR product arrays and oligonucleotide arrays which are free of other DNAs. Thus, the clone arrays should have the highest background among the current choices and represent the greatest challenge for the probes we developed.

RNA fingerprinting for probe preparation

RAP-PCR amplifications were performed to look for differential gene expression in keratinocytes (HaCaT) (7) when treated with EGF or TGF-β for 4 h. Using RAP-PCR, ~1% of the genes in normal or immortal keratinocytes responded to EGF and fewer responded to TGF-β in this time frame (data not shown). Two fingerprints were chosen for hybridization to cDNA arrays. Figure 1A and B shows RAP-PCR fingerprints of RNA from confluent keratinocytes treated with TGF-β or EGF, using multiple RNA concentrations and two sets of arbitrarily chosen primers. Primarily, the untreated control and EGF-treated samples were further explored in this study. In the first fingerprint (Fig. 1A), two differentially amplified products were detected, which had been cloned and sequenced in the course of our previous work. The sizes of these two products are indicated (291 and 317 nt). This fingerprint was used to demonstrate that we could identify differentially regulated genes in an array without isolating, cloning and sequencing the RAP-PCR products. This fingerprint and the second fingerprint in Figure 1, which displayed no differential regulation in response to the treatments, were also used to demonstrate that fainter differentially regulated products not visible on the fingerprint gel could, nevertheless, be observed by the array approach.

The fingerprints in Figure 1 fulfill important criteria of reproducibility. To be suitable for either gel- or array-based analysis, RAP-PCR fingerprints must remain almost identical over an 8-fold dilution of the input RNA. Low quality RAP-PCR fingerprints are usually the consequence of poor control over RNA quality and concentration. Before proceeding with the array hybridization steps, it is wise to verify the high quality of the RAP-PCR step. Because the array method has such high throughput, this extra step is neither costly nor time consuming and can greatly improve efficiency by reducing the number of false positives due to poor fingerprint reproducibility.

RAP-PCR fingerprints chosen from Figure 1 were converted into high specific activity radioactive probes by random primed synthesis using [α-^{32}P]dCTP. For each of the two conditions, EGF-treated and untreated, fingerprints generated from RNA at two different concentrations were converted to probe by random primed synthesis for each of the two different fingerprinting

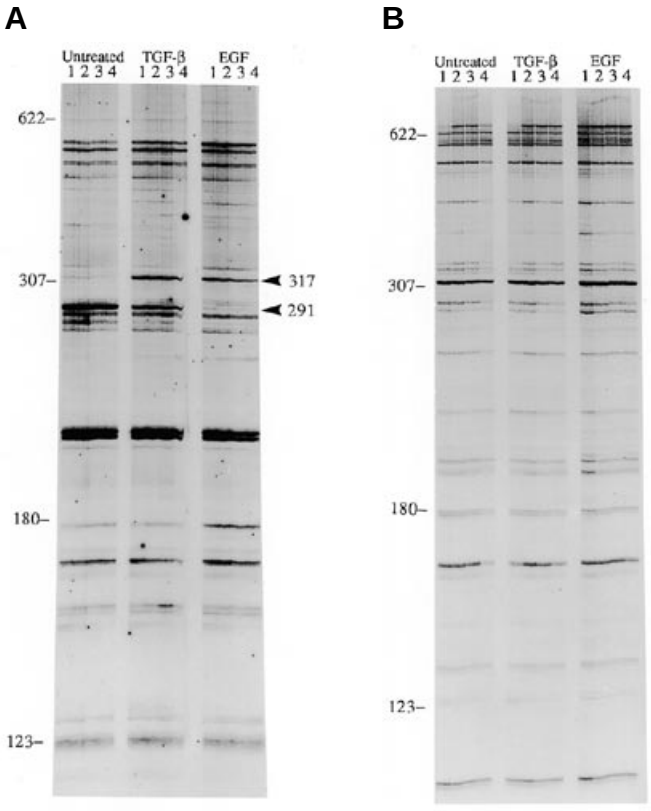


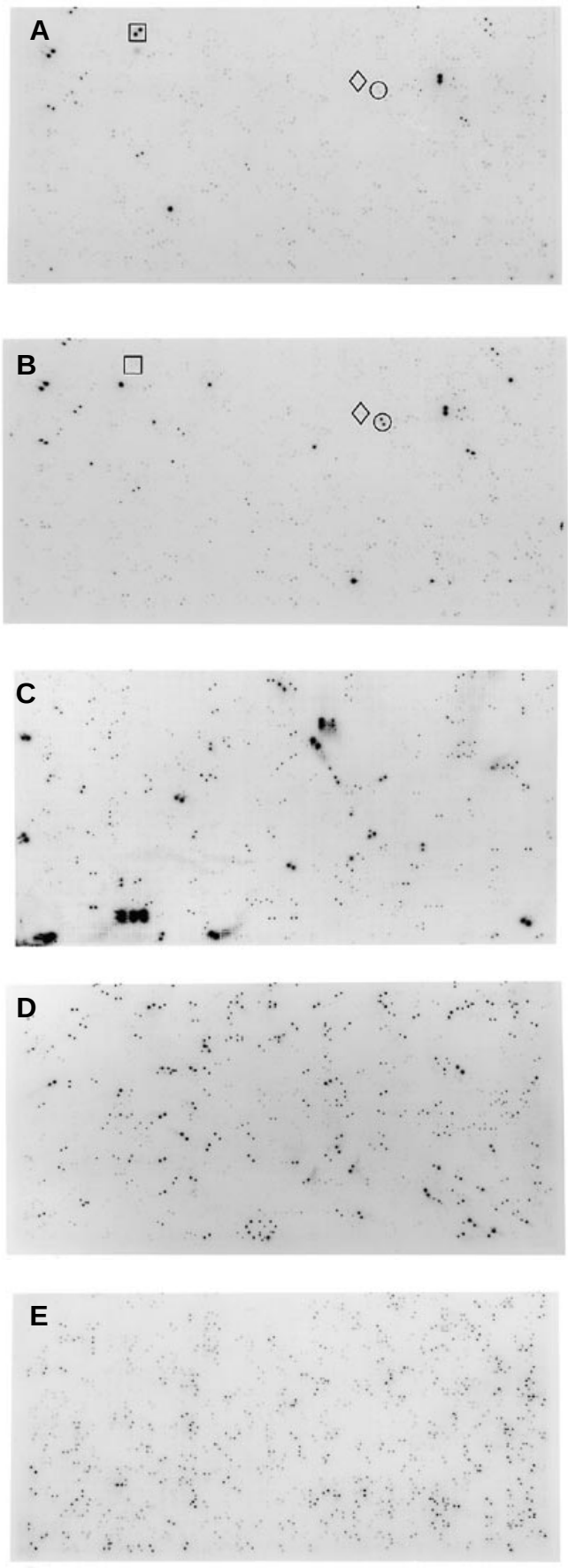
Figure 1. RAP-PCR fingerprints resolved on a gel. Reverse transcription was performed with an oligo(dT) primer on 250, 125, 62.5 and 31.25 ng RNA in lanes 1–4 respectively. RNA was from untreated, TGF-β- and EGF-treated HaCaT cells. RAP-PCR was performed with two sets of primers: (**A**) GP14/GP16; (**B**) Nuc1+/OPN24. Molecular weight markers and the sizes of the two differentially amplified RAP–PCR products are indicated.

primer pairs. These radioactively labeled probes were then hybridized to a set of identical arrays each containing 18 432 IMAGE consortium cDNA clones. As controls, total genomic DNA and total poly(A)$^+$ mRNA were also labeled by random priming and used as probes on identical arrays.

Hybridization to arrays

The probes derived from the RAP-PCR fingerprinting reactions described above and the total mRNA and genomic probes were used individually against replicates of a Genome Systems colony array. Hybridization and washing followed standard procedures outlined in Materials and Methods, including the use of genomic DNA as a blocking agent and as a competitor for highly repetitive sequences. Washing at 68°C in 0.1× SSC, 0.1% SDS removed virtually all hybridization to known Alu elements on the membrane, presumably because Alu elements are sufficiently diverged from one another at this wash stringency.

Autoradiograms from the same half of each membrane are shown in Figure 2. Data can also be collected using a phosphorimager, which considerably shortens data collection time and allows quantitation. Other means of labeling, such as fluorescently tagged bases, can be used if suitable arrays and instruments are available. Nylon membranes are typically unsuitable for most commercially



available fluorescent tags due to background fluorescence from the membrane itself.

Overlaps between different probes

The data were analyzed in a number of ways. First, estimates were made of the overlap between the clones hybridized by each probe. In all pairwise comparisons between all of the different types of probes, there was <5% overlap among the 500 clones that hybridized most intensely (compare Fig. 2A, B, D and E). Of the top 500 clones hybridized by the genomic probe (which included nearly all clones known to contain the Alu repeats), <5% overlapped with the top 500 clones hybridized by the fingerprint probes or the total poly(A)$^+$ mRNA probe. This indicated that, except for the case of genomic probe, there was no significant hybridization to dispersed repeats. The overlap among the clones hybridized by the two RAP-PCR fingerprints generated with different primers was <3% and the overlaps of either fingerprint with the poly(A)$^+$ mRNA probe were both <3%. Thus, most of the cDNAs detected using probe from the fingerprints could not be detected using the total mRNA probe. These data indicate that RAP-PCR samples a population of mRNAs largely independently of message abundance. This makes sense because the low abundance class of messages has much higher complexity than the abundant class, making it more likely that the arbitrary primers will find good matches. Unlike differential display, RAP-PCR demands two such arbitrary priming events, possibly biasing RAP-PCR toward the complex class. Overall, these data suggest that the majority of the mRNA population in a cell (<20 000 mRNAs) may be found in as few as 10 RAP-PCR fingerprints.

Further aspects of the data address reproducibility concerns. Using gel electrophoresis, there were no differences among the ~100 bands visible in any of the fingerprints from a single treatment condition performed at different RNA concentrations (Fig. 1). Similarly, >99% of the top 1000 clones hybridized by the probes derived from the fingerprint in Figure 1A were visible at both input RNA concentrations. Furthermore, >98% of the products were the same between the two treatment conditions (i.e. plus and minus EGF) at a single RNA concentration. This indicated almost perfect reproducibility among the top 1000 PCR products in the RAP-PCR amplification.

Figure 2. Hybridization to arrays. All images presented are autoradiograms of the bottom half of duplicates of the same Genome Systems filter probed with radiolabeled DNA. (**A**) and (**B**) Two RAP-PCR reactions using the same primers; (A) untreated; (B) EGF-treated. Three double-spotted clones that show differential hybridization signals are marked on each array. The GenBank accession nos of the clone and the corresponding genes are: square, H10045 and H10098, *vav-3*, AF067817 (13); circle, H28735, gene unknown, similar to heparan sulfate 3-*O*-sulfotransferase-1, AF019386 (17); diamond, R48633, gene unknown. A >10-fold down-regulation for *vav-3*, a >10-fold up-regulation for H28735 and an ~3-fold up-regulation of R48633 were independently confirmed by RT–PCR. (**C**) RAP-PCR using the same RNA as in (A) but with a different pair of primers yields an entirely different pattern. (**D**) cDNA, generated by reverse transcription of 1 µg poly(A)$^+$-selected mRNA. (**E**) Human genomic DNA labeled using random priming.

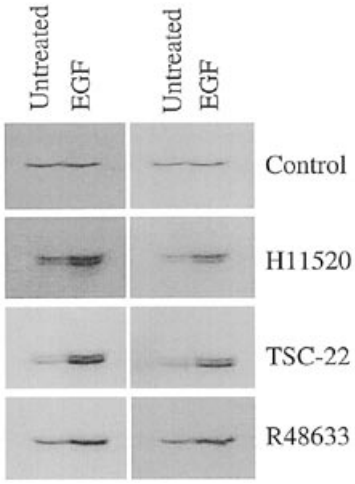


Figure 3. Confirmation of differential regulation by low stringency RT–PCR. Reverse transcription was performed at two RNA concentrations (500 ng, left column; 250 ng, right column). The reaction was diluted 4-fold in water and one fourth was used for low stringency RT–PCR at different cycle numbers. Shown are the control bands, the bands for GenBank accession no. H11520 (both at 22 cycles), the bands for TSC-22 [H11073 and H11161 (27–29)] and the bands for R48633 (all at 19 cycles). H11520 and TSC-22 are ~8–10-fold up-regulated by EGF. R48633 is ~3-fold up-regulated.

The detection of differentially regulated genes using RAP-PCR-derived probes against cDNA arrays

These experiments were designed to detect genes differentially regulated by EGF and TGF-β treatment in confluent keratinocytes. The fingerprint in Figure 1A reveals two boldly differentially regulated genes, the sequences of which were determined during the course of previous work (data not shown). The choice of which Genome Systems arrays to use was based on the presence of these clones. Figure 2 shows the results of hybridization of probes from these fingerprints to the arrays. Arrayed clones corresponding to the 291 nt (*vav-3*, square) and 317 nt (similar to N-HSST, circle) sequenced RAP-PCR fragments are indicated (compare Fig. 2A and B).

Also indicated on this array is a differentially regulated gene that could not be visualized on the original fingerprint gel. This result indicates that differential gene regulation can be detected by the combined fingerprinting and array approach even when the event cannot be detected using the standard gel electrophoresis approach. Verification of differential expression was performed by RT–PCR and will be described in the next section.

A total of 30 differentially hybridizing cDNA clones were detected among ~2000 hybridizing colonies using probes derived from both sets of arbitrary primers (Fig. 1) at a threshold of ~3-fold differential hybridization. Twenty two of these differentially hybridizing clones displayed differential hybridization at both RNA concentrations. These 22 were carried further to the RT–PCR confirmation step, described in the next section.

The eight false-positive clones that appeared to be regulated at only one concentration were of interest in exploring sources of error in the system. Of these eight, five potentially miscalled cases showed differential hybridization at one concentration but were present and not regulated on the membranes for the other concentration. The most likely source of this type of error is in the membranes. Although each clone is spotted twice, it is possible that occasionally one membrane received substantially more (or less) DNA in both spots than the other three membranes for these clones. However, this potential error was easily detected and is rare, occurring only five times in >2000 clones. The other three potentially miscalled cases hybridized under only one treatment condition and at only one RNA concentration used for RAP-PCR. These may be real differentially expressed genes, but might be false positives from irreproducible PCR products. However, there is an extraordinarily low number of these irreproducible products in the experiments we present here and they are easily identified by comparing the results of two probes derived from PCR of different starting concentrations of RNA.

Confirmation of differential expression using low stringency RT–PCR

Only those hybridization events that indicated differential expression at both input RNA concentrations were carried further. For confirmation of differential expression, we used RT–PCR with specific probes rather than northern blots because we expected that many of the mRNAs would be rare and northern blots are much less sensitive than RT–PCR. One of the advantages of using the arrays from the IMAGE consortium is that >70% of the clones have single pass sequence reads from the 5′- or 3′-end or both deposited in the GenBank database. Thus, it is usually not necessary to sequence a clone in order to derive primers for specific PCR. In cases where there is no sequence available, the clones can be ordered from Research Genetics and sequenced. We have used this strategy in the past, but in this report we confine ourselves to clones for which some sequence is available in the database. Five of the 22 ESTs representing differentially regulated genes on the array had not been sequenced and two of the remaining 17 ESTs were from the same gene. This left 15 unique sequenced genes. In all cases we attempted to align sequences from differentially regulated genes with other sequences in the database in order to derive a higher quality sequence from multiple reads and longer sequence from overlapping clones. The UniGene database clusters human and mouse ESTs that appear to be from the same gene (10). This database greatly aids in the process of assembling a composite sequence from different clones of the same mRNA (http:// www.ncbi.nlm.nih.gov/UniGene/ index.html). These composite sequences were then used to choose primers for RT–PCR.

For each gene, two specific primers were used in RT–PCR under low stringency conditions similar to those used to generate RAP-PCR fingerprints. In addition to the product of interest, a pattern of arbitrary products is generated which is largely invariant and behaves as an internal control for RNA quality and quantity and for reverse transcription efficiency (11). The number of PCR cycles was adjusted to between 14 and 25 cycles, according to the abundance of the product, in order to preserve the differences in starting template mRNA abundances. This is necessary because rehybridization of abundant products during the PCR inhibits their amplification and the difference in product abundances diminishes as the number of PCR cycles increases, in what we have called the ‘Cot effect’ (12).

Low stringency RT–PCR experiments confirmed the differential expression of the two transcripts that were identified in the RAP-PCR fingerprints of Figure 1A and showed differential

hybridization to the cDNA array (Fig. 2A versus B). These genes had previously been isolated from the gel in Figure 1 and sequenced. One of these corresponds to a new family member of the *vav* proto-oncogene family (13–16) and the other has homology to heparan sulfate 3-*O*-sulfotransferase-1 (17). These have been shown to be regulated under a variety of experimental conditions (manuscript in preparation). The other 13 candidates were also tested and 11 were confirmed. Examples are presented in Figure 3. A list of these genes is given in Table 1. Of the two that were not confirmed, one proved unamplifiable, perhaps because of the low quality sequence used to make the primers or because hybridization to the array was by a differentially regulated closely similar family member. The other gene gave a product but appeared not be differentially regulated by RT–PCR. In addition to the possibility of a family member being regulated, this result could also be due to differential processing of the mRNA rather than differential promotor activity. There is already a precedent for this: differential processing appears to be the reason that *vav-3* yields differential hybridization and differential processing is only observed if the correct primers are chosen (manuscript in preparation).

Table 1. Genes regulated by >2-fold after EGF treatment of confluent HaCaT keratinocytes[a]

Accession number	Gene name
Up-regulated	
H11520 (3′)	Unknown
H11161 (5′)/H11073 (3′)	TSC-22 [U35048]
R48633 (5′)	Unknown
H28735 (3′)	Similar to heparan sulfate 3-*O*-sulfotransferase-1 precursor [AF019386]
H25513 (5′)/H25514 (3′)	Fibronectin receptor α subunit [M13918]
H12999 (5′)/H05639 (3′)	Similar to focal adhesion kinase (FAK2) [L49207]
H15184 (5′)/H15124 (3′)	*ray* gene [X79781]
H25195 (5′)/H24377 (3′)	X-box binding protein-1 (XBP-1) [M31627]
H23972 (3′)	Unknown
H27350 (5′)	CPE receptor (hCPE-R) [AB000712]
R75916 (5′)	Similar to semaphorin C [X85992]
Down-regulated	
R73021 (5′)/R73022 (3′)	Epithelium-restricted Ets protein ESX [U66894]
H10098 (5′)/H10045 (3′)	*vav-3* [AF067817]

[a]Differential expression was confirmed by low stringency RT–PCR. The left column gives the accession numbers of the EST clones (5′ or 3′ or both when available). The right column gives the corresponding gene or the closest homolog. In cases of very low homlogies the gene is considered unknown.

Detecting rare mRNAs

How effective are RAP-PCR probes at detecting rarer mRNAs? Each fingerprint hybridizes to a set of clones almost entirely different from the set hybridized by a probe derived from poly(A)$^+$-selected mRNA (Fig. 2). In addition, numerous other primer pairs, membranes and sources of RNA consistently show a <5% overlap between clones hybridized by any two fingerprints or between a fingerprint and a total poly(A)-selected cDNA probe (data not shown). We also attempted to use a northern blot of poly(A)-selected RNA to detect the *vav-3* mRNA (Fig. 1A), which is a new member of the *vav* oncogene family. Despite our ability to detect serially diluted vector down to the equivalent of a few copies per cell, we were unable to detect *vav-3* mRNA, whereas RT–PCR confirmed expression. A glyceraldehyde 3-phosphate dehydrogenase control indicated that the northern blot was performing correctly (data not shown). *vav-3*, therefore, appears to be a low abundance message that is represented in a RAP-PCR fingerprint as a prominent band.

The frequency of homologs of cDNAs detected by the RAP-PCR probes in the EST database was determined (>98% identity). This was compared with the frequency of homologs for a random set of other cDNAs on the same membrane. If the RAP-PCR fingerprints were heavily biased towards common mRNAs, then many would occur often in the EST database because it is partly derived from cDNA libraries that are un-normalized or incompletely normalized. However, the cDNAs detected by RAP-PCR had frequencies in the EST database comparable with the frequencies for randomly selected cDNAs, including cases where the clone was unique in the database. This implies that sampling by RAP-PCR is at least as good as random sampling of the partly normalized libraries used to construct the array, and certainly very different from that obtained for an un-normalized probe such as total mRNA.

Comparison with other sampling methods

In principle, there are several ways to generate a reduced complexity cDNA probe. One of the most successful ways to reduce probe complexity while accentuating the differences between two samples is to perform subtraction (see for example 18), which can have a sensitivity of 1/200 000 (19). It is an obvious but important extension of this manuscript that it would be worthwhile to screen mixtures resulting from subtraction using arrays of ESTs or total cDNAs, when they become available.

Subtraction can be applied to RAP-PCR by simply quenching a labeled fingerprint with an unlabeled fingerprint and we have preliminary evidence that this works (data not presented). A limitation of subtraction is that it can eliminate differences that fall short of presence versus total absence of a mRNA. Furthermore, while subtraction is useful in a binary question, it is of limited utility in cases where a large number of conditions are to be compared combinatorially.

There are two fundamentally different types of complexity reduction: those that maintain the relative stoichiometry among the mRNAs they sample, and those which do not. In the former category are strategies such as selecting a narrow size class of mRNAs or cDNAs (20), where rare mRNAs would still be rare. Other methods that maintain approximate stoichiometry include those that employ 3′-anchored cDNA restriction fragments (see for example 21–23). In a RAP-PCR both the abundance and the degree of match with the primers contribute to the prevalence of any particular product. Thus, rare mRNAs that happen to have excellent matches with the primers and are efficiently amplified are found among the more abundant RAP-PCR products. In this respect, a RAP-PCR probe is non-stoichiometric (24). This is a

very useful feature of RAP-PCR because it allows the sampling of mRNAs that are difficult to sample using other methods.

Changing the number of products sampled by RAP-PCR

Detection is ultimately limited by background hybridization and incomplete blockage of repeats. At present, ~1000 cDNAs on a colony array of 18 432 clones can be reliably scored by each RAP-PCR probe and the limitation seems to be the number of sufficiently abundant products in the PCR reaction rather than background.

The effect of RAP-PCR reaction parameters on the distribution and number of products that can be observed on arrays has not been fully explored, including the optimal complexity of the probe. To increase the complexity we used *Taq* polymerase Stoffel fragment, which is more promiscuous than Ampli*Taq*. The primers used were 10 or 11 bases in length and are not degenerate (they have a single base at each position). Longer primers used at the same temperature might give a more complex product, as would primers with some degeneracy. We have recently used an oligo(dT) primer anchored at the 3′-end as one of the two primers (manuscript in preparation). Anchoring at the 3′-end of messages (25) should result in more hybridization in arrays that are 3′-biased. However, the greater the complexity of the probe, the more closely it will resemble a total mRNA probe, which loses the advantage of non-stoichiometric sampling. Using arrays will teach us a number of things about the RAP-PCR mixtures that would be much less evident from a gel. For example: How complex are the fingerprints? What is the distribution of products among those easily seen on a gel and those that are too rare to be seen on a gel? What are the various effects of primer length, degeneracy and anchoring in the reverse transcription and PCR reactions? What are the effects of various different polymerases at each of these steps? Some of the answers to these questions will undoubtedly improve the throughput of the method for arrays.

Colony arrays used in these experiments represent the worst case scenario, in which plasmid DNA is mixed with a large mass excess of a bacterial genome having 5 Mb of complexity. If probes become so complex that background becomes the limiting factor, more sophisticated arrays may become essential. PCR product arrays or oligonucleotide arrays may yield higher scorability with more complex probes.

Comparison with gel-based characterization of fingerprints

Cloning genes from RAP-PCR fingerprints resolved on gels may still have some advantages over using fingerprints for probes on arrays in certain circumstances. For example, a new gene could be found that is not already on a membrane. However, this advantage diminishes every day as more cDNAs are characterized. Another advantage is that close family members can yield different PCR products in fingerprints, whereas on arrays a close family member may hybridize to a clone and lead to a misinterpreted result. This possibility is somewhat diminished if 3′ ESTs are used on membranes, because 3′-ends of even quite closely related genes may be sufficiently divergent to avoid that problem. Furthermore, fingerprints resolved on gels may detect new splicing variants, which is less likely using clone arrays. Also, when using gel-based fingerprints a single primer pair can be used to survey 100 mRNAs in a very large number of RNA samples. The same number of fingerprints applied to arrays would be expensive, though this is balanced by the fact that 1000 mRNAs would be surveyed. Overall, for most applications, the ability to screen many thousands of genes with a single fingerprint on a series of arrays will outweigh the advantages of gel-based assays. In addition, any bias toward abundant transcripts that exists in RAP-PCR is partially mitigated by the array approach, because even products that are never visible in the gels can still serve as effective probes.

The rate of throughput using fingerprints as probes for arrays compares favorably with that obtainable using gel fingerprints. A single sequencing-style gel loaded with RAP-PCR fingerprints from 25 different primer pairs usually displays ~1000 products. This is similar to the number of mRNAs we surveyed on a single membrane containing about one third of the unique sequences in the IMAGE consortium clones. A fingerprinting experiment using 25 primer pairs could supply probes for cDNA arrays conceivably covering >20 000 genes.

Hybridization of fingerprints to arrays has the huge advantage that there is generally no need to isolate, clone and sequence the genes detected. In principle, all known human mRNAs will fit on three membranes (~50 000 genes). At present, each fingerprint has sufficient complexity to hybridize to >2000 of the 50 000 known genes. There is also the issue of diminishing returns. In a fingerprint, one cannot know if a differentially amplified product has been sampled previously without performing considerable further work. In contrast, with an array one always knows what genes have been sampled previously. In principle, one can even select primers that enrich for genes not yet sampled (26).

In summary, a method is presented in which the intrinsic reduced complexity and non-stoichiometric amplification resulting from arbitrarily primed PCR fingerprinting is used to advantage to construct probes for cDNA arrays. Simple methods that allow inexpensive arrays to generate useful information are likely to allow many molecular biology laboratories to participate in the revolution in understanding gene regulation that arrays can achieve. We hope that a public resource will soon develop in which the transcriptional effect of a growing list of conditions is attached to every gene. Ultimately, such information will link to the promotors of these genes and to the signal transduction cascades responsible for their regulation.

ACKNOWLEDGEMENTS

This work was supported in part by a generous gift from Sidney Kimmel, NIH grants CA68822, NS33377, AI34829 to M.M., DOD grant BC961294 to J.W., DFG fellowships TR413/1-2 to T.T. and JU355/1-2 to B.J. and a TRDRP grant 6KT-0272 to F.M.-D. We thank Rhonda Honeycutt for critical reading of the manuscript.

REFERENCES

1 Marshall,A. and Hodgson,J. (1998) *Nature Biotechnol.*, **16**, 27–31.
2 Ramsay,G. (1998) *Nature Biotechnol.*, **16**, 40–44.
3 Pietu,G., Alibert,O., Guichard,V., Lamy,B., Bois,F., Leroy,E., Mariage-Sampson,R., Houlgatte,R., Soularue,P. and Auffray,C. (1996) *Genome Res.*, **6**, 492–503.
4 Boll,W., Fujisawa,J., Niemi,J. and Weissmann,C. (1986) *Gene*, **50**, 41–53.
5 Welsh,J. and McClelland,M. (1990) *Nucleic Acids Res.*, **18**, 7213–7218.
6 Welsh,J., Chada,K., Dalal,S.S., Cheng,R., Ralph,D. and McClelland,M. (1992) *Nucleic Acids Res.*, **20**, 4965–4970.

7 Boukamp,P., Popp,S., Altmeyer,S., Hulsen,A., Fasching,C., Cremer,T. and Fusenig,N.E. (1997) *Genes Chromosomes Cancer*, **19**, 201–214.
8 McClelland,M., Ralph,D., Cheng,R. and Welsh,J. (1994) *Nucleic Acids Res*., **22**, 4419–4431.
9 Mathieu-Daude,F., Trenkle,T., Welsh,J., Jung,B., Vogt,T. and McClelland,M. (1998) *Methods Enzymol*., in press.
10 Schuler,G.D. (1997) *J. Mol. Med*., **75**, 694–698.
11 Mathieu-Daude,F., Welsh,J., Davis,C. and McClelland,M. (1998) *Mol. Biochem. Parasitol*., **92**, 15–28.
12 Mathieu-Daude,F., Welsh,J., Vogt,T. and McClelland,M. (1996) *Nucleic Acids Res*., **24**, 2080–2086.
13 Katzav,S., Martin-Zanca,D. and Barbacid,M. (1989) *EMBO J*., **8**, 2283–2290.
14 Katzav,S. (1995) *Crit. Rev. Oncogen*., **6**, 87–97.
15 Bustelo,X.R. (1996) *Crit. Rev. Oncogen*., **7**, 65–88.
16 Romero,F. and Fischer,S. (1996) *Cell Signalling*, **8**, 545–553.
17 Shworak,N.W., Liu,J., Fritze,L.M., Schwartz,J.J., Zhang,L., Logeart,D. and Rosenberg,R.D. (1997) *J. Biol. Chem*., **272**, 28008–28019.
18 Jin,H., Cheng,X., Diatchenko,L., Siebert,P.D. and Huang,C.C. (1997) *BioTechniques*, **23**, 1084–1086.
19 Rhyner,T.A., Biguet,N.F., Berrard,S., Borbely,A.A. and Mallet,J. (1986) *J. Neurosci. Res*., **16**, 167–181.
20 Dittmar,G., Schmidt,G., Kopun,M. and Werner,D. (1997) *Cell Biol. Int*., **21**, 383–391.
21 Bachem,C.W., van der Hoeven,R.S., de Bruijn,S.M., Vreugdenhil,D., Zabeau,M. and Visser,R.G. (1996) *Plant J*., **9**, 745–753.
22 Habu,Y., Fukada-Tanaka,S., Hisatomi,Y. and Iida,S. (1997) *Biochem. Biophys. Res. Commun*., **234**, 516–521.
23 Money,T., Reader,S., Qu,L.J., Dunford,R.P. and Moore,G. (1996) *Nucleic Acids Res*., **24**, 2616–2617.
24 Trenkle,T., Mathieu-Daude,F., Welsh,J. and McClelland,M. (1998) *Methods Enzymol*., in press.
25 Liang,P. and Pardee,A.B. (1992) *Science*, **257**, 967–971.
26 Pesole,G., Liuni,S., Grillo,G., Belichard,P., Trenkle,T., Welsh,J. and McClelland,M. (1998) *BioTechniques*, **25**, 112–117.
27 Jay,P., Ji,J.W., Marsollier,C., Taviaux,S., Berge-Lefranc,J.L. and Berta,P. (1996) *Biochem. Biophys. Res. Commun*., **222**, 821–826.
28 Dmitrenko,V.V., Garifulin,O.M., Shostak,E.A., Smikodub,A.I. and Kavsan,V.M. (1996) *Tsitol. Genet*., **30**, 41–47.
29 Ohta,S., Shimekake,Y. and Nagata,K. (1996) *Eur. J. Biochem*., **242**, 460–466.