5640–5648 Nucleic Acids Research, 1994, Vol. 22, No. 25

Identifying differences in mRNA expression by representational difference analysis of cDNA

M.Hubank and D.G.Schatz*
Section of Immunobiology, Howard Hughes Medical Institute, Yale University School of Medicine, 310 Cedar Street, Box 208011, New Haven, CT 06520-8011, USA

Received September 2, 1994; Revised and Accepted November 12, 1994

ABSTRACT

Detection of differentially regulated genes has been severely hampered by technical limitations. In an effort to overcome these problems, the PCR-coupled subtractive process of representational difference analysis (RDA) [Lisitsyn,N. *et al.* (1993) Science 259, 946–951] has been adapted for use with cDNA. In a model system, RAG-1 and RAG-2, the genes responsible for activating V(D)J recombination, were identified in a genomic transfectant by cDNA RDA in a small fraction of the time taken by conventional means. The system was also modified to eliminate expected difference products to facilitate the identification of novel genes. Additional alterations to the conditions allowed isolation of differentially expressed fragments. Several caffeine up-regulated clones were obtained from the pre-B cell line 1-8, including IGF-1B, and a predicted homologue of the natural killer cell antigen, NKR-P1. The approach was found to be fast, extremely sensitive, reproducible, and predominantly lacked false positives. cDNA RDA has the capacity and adaptability to be applied to a wide range of biological problems, including the study of single gene disorders, characterization of mutant and complemented cell types, developmental or post-event expression time courses, and examination of pathogen–host interactions.

INTRODUCTION

Identifying changes in gene expression is a frequent goal of modern biomedical research. Differential gene expression separates differentiated cells from stem cells, activated cells from their resting counterparts and mutant from normal phenotypes. Determining the differences that exist in the mRNA complement between these otherwise closely related groups contributes information essential for unraveling the biochemical complexity of metabolic events. A variety of methods are currently employed to isolate gene products associated with particular phenotypes. Subtractive hybridization of cDNA with mRNA, or of cDNA libraries, has been used to identify messages responsible for certain developmental events, but is technically difficult, time consuming and often either impractical or unreliable (1). Similar reservations apply to complementation of mutants with cDNA libraries or with genomic DNA. The ability to rapidly reduce the number of candidate genes to a few which could be easily characterized would be greatly advantageous.

Recently two techniques with this potential have been described. Both employ PCR to amplify messages to detectable levels, but their mode of operation is fundamentally different. Differential display PCR (DDPCR) relies on random-primed amplification of a subfraction of total mRNA from two populations, running the amplicons side by side on a sequencing gel, and isolating bands which are expressed at different levels (2). Representational difference analysis (RDA) is a process of subtraction coupled to amplification, originally developed for use with genomic DNA as a method capable of isolating the differences between two complex genomes (3). Differential display amplifies fragments from all represented mRNA species, whereas RDA eliminates those fragments present in both populations, leaving only the differences. For this reason, we decided to adapt RDA for use with cDNA.

Genomic RDA relies on the generation, by restriction enzyme digestion and PCR amplification, of simplified versions of the genomes under investigation known as 'representations'. If an amplifiable restriction fragment (the target) exists in one representation (the tester), and is absent from another (the driver), a kinetic enrichment of the target can be achieved by subtractive hybridization of the tester in the presence of excess driver (3). Sequences with homologues in the driver are rendered unamplifiable, while the target hybridizes only to itself, and retains the ability to be amplified by PCR. Successive iterations of the subtraction/PCR process produce ethidium visible bands on an agarose gel corresponding to enriched target (3).

It is necessary to reduce the complexity of the starting material for genomic RDA. This is achieved by digesting total genomic DNA with a six-cutting enzyme, and amplifying the digest by PCR. A high proportion of the digested fragments do not fall into the amplifiable range of 150–1000 base pairs, reducing the complexity of the amplicon so that the final representation contains only about 2–10% of the total genome (3,4). In contrast, a population of cDNA derives from some 15 000 different genes in a typical cell, representing only about 1–2% of the total

*To whom correspondence should be addressed

genome (5). Sufficiently few sequences are present in cDNA that RDA can be applied without the need to reduce its complexity. We therefore restricted cDNA with a four-cutting enzyme (*Dpn*II) in order to generate the representations. A mean cutting length of approximately 256 base pairs ensures that the vast majority of cDNA species will contain at least one amplifiable fragment—sufficient to isolate a difference and identify the gene.

The ability of cDNA RDA to detect absolute differences between two populations offers a number of immediate applications. Mutants resulting from the lack of expression of an unknown gene and transfectants which alter cell phenotypes may both be readily characterized in this way. In many cases, however, gene expression may not be absolutely affected after an event, but may vary in scale. The stringency of hybridization employed in standard RDA, in terms of amount of driver necessary to gain sufficient target enrichment, renders these differences invisible. We have further explored the capacity of the system, and found that by modification of the tester:driver ratio, it is possible to bias the kinetic enrichment in favor of species up-regulated from basal levels.

In all cases cDNA RDA has been assessed in real biological situations. Test systems were provided by studying various aspects of the recombination activation genes (RAG-1 and RAG-2), involved in the site-specific V(D)J (variable, diversity, joining) recombination process which assembles immunoglobulin and T cell receptor genes. We were able to isolate the RAGs from a transfected cell line in a small fraction of the time originally required to do so. We were also successful in competing out unwanted difference product from the reactions, and finally employed the modified technique to identify a number of caffeine induced cDNA fragments from a pre-B cell line. In the course of these studies we also found that cDNA RDA is able to identify genes expressed in only a very small fraction (less than 1%) of the cells making up the tester population, demonstrating that the technique is substantially more sensitive than any other subtractive cDNA method.

MATERIALS AND METHODS

Oligonucleotides

Sequences of oligonucleotides used in cDNA RDA were as follows: R-Bgl-24 5′-AGCACTCTCCAGCCTCTCACCGCA-3′, R-Bgl-12 5′-GATCTGCGGTGA-3′; J-Bgl-12 5′-GATCTGTTCATG-3′, J-Bgl-24 5′-ACCGACGTCGACTATCCATGAACA-3′; N-Bgl-12 5′-GATCTTCCCTCG-3′; N-Bgl-24 5′-AGGCAACTGTGCTATCCGAGGGAA-3′ (3).

General methods

All common techniques and routine DNA manipulations, including transformations, plasmid preparations, cesium chloride gradients and gel electrophoresis, were carried out according to standard procedures (6). Restriction and modifying enzymes (Boehringer Mannheim and New England Biolabs) were used in accordance with manufacturer's recommendations. For northern blots, cytoplasmic RNA was fractionated on formaldehyde agarose gels and transferred onto GeneScreen Plus nylon membrane (New England Nuclear), following the manufacturer-supplied protocols. Probes were labeled by random primer extension in the presence of [^{32}P]-dCTP (>3000 Ci/mM, Amersham International), and hybridizations carried out in 50% formamide, 2.5×Denhardt's solution, 1% SDS, 4×SSPE, 250 μg/ml salmon sperm DNA (18 h, 42°C). Membranes were washed twice in 0.1% SDS, 2×SSC (10 min, 25°C), followed by once each in 0.1% SDS, 2×SSC (15 min, 65°C); 0.1% SDS, 0.5×SSC (15 min, 65°C); and 0.1% SDS, 0.25×SSC (15 min, 65°C), before autoradiography.

Cytoplasmic RNA was prepared by NP40 cell lysis, removal of nuclei by centrifugation, and polyA$^+$ RNA was isolated by twice selecting over an oligo-dT cellulose column (7). Double stranded cDNA was prepared by reverse transcription of the polyA$^+$ RNA using an oligo dT primer (7), and its quality assessed on an agarose gel.

Cell culture and retrovirus infection

The pre-B cell line 204-1-8 (8), referred to throughout this paper as 1-8, was grown in RPMI 1640 medium supplemented with 10% fetal calf serum and 50 μM β-mercaptoethanol. For caffeine induction, 1-8 cells were grown with 1 mM caffeine for 12 h, after which the cells were lysed, and poly A$^+$ mRNA rapidly isolated. NIH3T3 and 3TGHR cells were grown in DMEM supplemented with 10% calf serum. The NIH3T3 fibroblast cell lines 3TGHR and TXH8 were derived from the cell line 3TGR (9): TXH8 is a population of DHR-infected TRX-1 cells selected in hygromycin-B [see Figure 4 of (9)] which stably expresses RAG-1 and RAG-2 and hence V(D)J recombinase activity, while 3TGHR is a population of 3TGR cells in which approximately 5−10% of the cells are infected with the DHR retrovirus (9). TXH8 cells were grown in DMEM lacking histidine supplemented with 500 μM histidinol, after confirming that the cells were resistant to both mycophenolic acid and hygromycin B (9).

The RAG-1 and RAG-2 expressing ecotropic retroviruses used to infect NIH3T3 cells were pMFG-RAG-1 and pMFG-RAG-2, (kind gifts of D.Silver, M.Sadelain, D.Baltimore and R.Mulligan, Whitehead Institute). The β-galactosidase expressing retrovirus was pMFGβgal (kind gift of M.Sadelain and R.Mulligan). Retroviral producer lines were cultured in DME containing 10% calf serum, supplemented with G418 (1 mg/ml). Virus-containing supernatants were harvested and applied to NIH3T3 cells in the presence of polybrene (8 μg/ml). Cells were incubated with the virus (37°C, 4 h), then diluted with 10 ml DME (10 % FCS), and incubated overnight at 37°C, after which the virus was removed and standard medium returned. Cells were subsequently harvested and DNA and RNA prepared. Southern and Northern blots confirmed the presence of high levels of both retroviral DNA and transcripts in the infected cells. βgal infected cells were stained with Xgal (37°C, 2−6 h) to determine infection efficiency.

cDNA RDA: generation of representations

Representation difference analysis of cDNA is an extension of the technique applied to genomic DNA (3) and is based on a protocol supplied by N.Lisitsyn (Cold Spring Harbor) (Figure 1). A detailed protocol is available from the authors on request. Double stranded cDNA (2 μg) was digested with *Dpn*II, phenol extracted, ethanol precipitated and resuspended in 20 μl TE. Twelve μl (~1.2 μg) of cut cDNA was then ligated to the R-Bgl-12/24 adapter in a mixture containing: 4 μl desalted R-Bgl-24 oligo (2 mg/ml), 4 μl desalted R-Bgl-12 oligo (1 mg/ml), 6 μl 10×ligase buffer (New England Biolabs), and 31 μl water. Oligonucleotides were annealed to each other and to the cDNA in a PCR machine by heating the ligation reaction to 50°C, then

cooling to 10°C over a period of 1 h, and ligation carried out by adding 3 μl T4 DNA ligase (400 U/μl), and incubating for 12–14 h at 12–16°C. Ligations were diluted to 6 μg/ml, and multiple PCR reactions were set up to generate the initial representations. Each 200 μl reaction contained 2 μl diluted ligation and (final concentrations) 66 mM Tris–HCl, pH 8.8 at 25°C; 4 mM μgCl_2; 16 mM $(NH_4)_2SO_4$; 33 μg/ml BSA; dATP, dCTP, dGTP and TTP (all 0.3 mM) and 2 μg R-Bgl-24 primer. The 12mer was melted away (3 min, 72°C) and the 3′ ends were filled in with 5 U of Taq DNA polymerase (BRL) (5 min, 72°C). Twenty cycles of amplification were performed (1 min, 95°C; 3 min, 72°C), and the products were combined, phenol extracted, ethanol precipitated and resuspended in TE at 0.5 μg/μl to provide the representation. The R-adapters were removed from the representations with *Dpn*II and the digest was phenol extracted and ethanol precipitated to form the driver. A portion of this digested representation (20 μg) was gel-purified on a 1.2% TAE agarose gel, and the products (free of the R-adapters) isolated using Qiaex resin (Qiagen). This formed the tester, of which 2 μg were ligated to the J-Bgl-12/24 adapter in the manner described above.

Hybridization and selective amplification

For the first subtractive hybridization, 0.4 μg (40 μl) J-ligated tester was mixed with 40 μg (80 μl) driver. The mixture was phenol extracted, ethanol precipitated, and resuspended in 4 μl EEx3 buffer [30 mM EPPS (Sigma), pH 8.0 at 20°C; 3 mM EDTA]. The solution was overlaid with mineral oil and DNA denatured (6 min; 98°C). The salt concentration was adjusted with 1 μl 5M NaCl and the sample was allowed to anneal (20 h, 67°C). The hybridized DNA was diluted with 8 μl TE containing 5 μg/μl yeast RNA, and thoroughly resuspended in 400 μl TE. For each subtraction, four 200 μl PCR reactions were set up as before with 20 μl diluted hybridization mix, but omitting the primer. The J-Bgl-12 oligonucleotide was melted away (3 min, 72°C), ends filled in with 5 U of Taq DNA polymerase (5 min, 72°C), and 2 μg J-Bgl-24 primer added. Ten cycles (1 min, 95°C; 3 min, 70°C) of amplification were performed, and the four reactions were combined, phenol extracted, isopropanol precipitated and resuspended in 40 μl 0.2×TE. Twenty μl of product was digested with 20 U (2 μl) mung bean nuclease in 1×digestion buffer (New England Biolabs) (35 min, 30°C), and the reaction stopped by the addition of 160 μl 50 mM Tris–HCl (pH 8.9). The digest was heated to 98°C (5 min) and chilled on ice. Final amplifications (four per hybridization) were set up on ice, including 20 μl MBN-treated product and 2 μl J-Bgl-24 (1 mg/ml). One μl (5 U) Taq DNA polymerase was added at 80°C, and 18 cycles performed (1 min, 95°C; 3 min, 70°C). Products were combined, phenol extracted, isopropanol precipitated and the pellet resuspended at 0.5 μg/μl, giving the first difference product (DP1). J-Adapters on DP1 were changed for N-Bgl-12/24 adapters, and the process of subtractive hybridization and selective amplification repeated to generate the second difference product (DP2). The procedure was as described, with the difference that in the second hybridization, 50 ng tester was mixed with 40 μg driver, and in the amplifications, annealing and extension were carried out at 72°C. To generate a third difference product (DP3), 100 pg J-ligated DP2 was mixed with 40 μg driver, and the process repeated, except that the final amplification was performed for 22 cycles (1 min, 95°C; 3 min, 70°C).

Melt depletion of representations

In some experiments, driver and tester representations were depleted of low copy sequences to detect messages present at a higher level in the tester than in the driver. This was achieved by resuspending digested double stranded cDNA (1.2 μg) ligated to R-oligos (but prior to amplification) in 4 μl 3xEE buffer. This cDNA was melted (6 min; 98°C), 1 μl NaCl (5 M) added, and allowed to anneal for 1 h (67°C). After this time, the sample was diluted on ice to 400 μl, and PCR reactions set up in the manner described for generating representations. In preliminary experiments, these conditions were found to significantly eliminate low copy sequences from the subsequently amplified population.

Cloning and sequencing of difference products

Final difference products were digested with *Dpn*II and cloned into the *Bam*HI site of pBluescript KS$^+$ II (Stratagene). Double stranded plasmid DNA was prepared using miniprep columns (Qiagen), and sequenced with an ABI Dyedeoxy Terminator Cycle Sequencing apparatus (Applied Biosystems). Resulting sequences were compared to the GenBank database using the BLAST program (10).

RESULTS

cDNA RDA detects low abundance transcripts from transfected fibroblasts

The cDNA RDA procedure described here is based closely on the genomic RDA procedure described by Lisitsyn *et al.* (3). RDA PCR amplification conditions limit amplification to fragments in the size range of 150–1000 bp. In the genomic RDA procedure, only 2–10% of genomic sequences are in this range because genomic DNA is digested with a 6-cutter restriction enzyme. This simplifies the amplicon but also excludes many sequences. We reasoned that because cDNA is already substantially less complex than the genome, such a simplification was unnecessary and therefore representations could be created using cDNA digested with a 4-cutter enzyme. As a consequence, the vast majority of cDNAs should yield at least one fragment in the amplifiable size range, and should therefore be able to be isolated by a single application of cDNA RDA. The method is shown schematically in Figure 1.

The cDNA RDA method was first applied to a test system consisting of two very similar NIH3T3 mouse fibroblast cell lines (TXH8 and 3TGHR) thought to differ only by the expression of several well characterized mRNAs. TXH8 is a genomic transfectant selected for its ability to perform V(D)J recombination. It expresses low levels of the recombination activating genes RAG-1 and RAG-2, as well as high levels of the co-transfected marker gene histidinol dehydrogenase (His) from the plasmid pSV2-His (9,11). 3TGHR, the untransfected control, was derived from the same founder cell line and should differ from TXH8 only in lacking expression of RAG-1, RAG-2 and His (and, in principle, any genes activated by expression of RAG-1 and RAG-2). To test the effectiveness of cDNA RDA, tester was prepared from TXH8 and driver from 3TGHR, and cDNA RDA was performed as described in Materials and Methods. A control experiment was performed in the reverse direction.

We saw a stepwise reduction of complexity of the products in each successive difference product, until clear bands with little

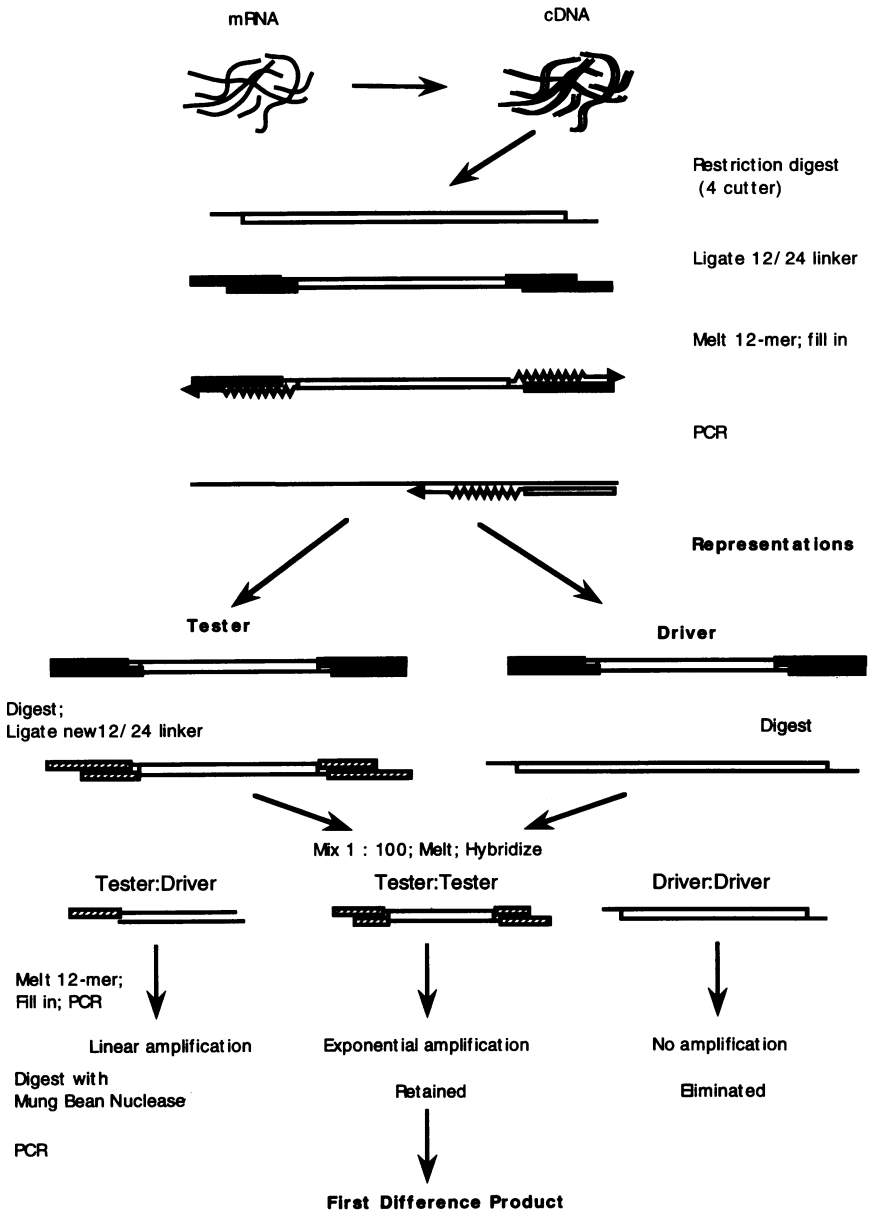

Figure 1. Schematic diagram of cDNA RDA, based on the genomic RDA technique of Lisitsyn *et al.* (3). Solid boxes represent the R-oligonucleotides used to generate the representations. Hatched boxes show the J (or N)-oligonucleotides used to generate difference products. Zig-zag arrows indicate DNA synthesis. The process is shown up to the first difference product. To generate second and third difference products, products are reintroduced into the scheme at the tester stage in the proportions indicated in Materials and Methods.

background were visible by ethidium staining in the third difference product (Figure 2A). The sizes of these bands corresponded to predicted amplifiable fragments of the cDNAs from RAG-1, RAG-2 and His. Difference products were transferred to membranes by Southern blotting, where hybridization to probes for RAG-1, RAG-2 and His accounted for all the major ethidium-visible species (Figure 2B, and data not shown).

High levels of tester message can be competed out by driver supplementation

It is likely that preferential amplification of particular PCR products may dominate a reaction, suppressing amplification of other genuine difference products. Similarly, the amplification of species already known to differ between two populations may impede the isolation of new ones. We therefore attempted to compete out dominant products by artificially supplementing the driver with amplifiable fragments known to be present in the tester, but absent from the driver.

We chose a test system in which it was hoped that the elimination of highly abundant RAG messages would allow the amplification of any species expressed as a result of RAG activity. NIH3T3 fibroblasts were infected with ecotropic retroviruses bearing the RAG genes (MFG-RAG-1 and MFG-RAG-2) or with an identical virus bearing the bacterial gene for β-galactosidase (MFG-βgal). DNA and RNA was prepared from the infected

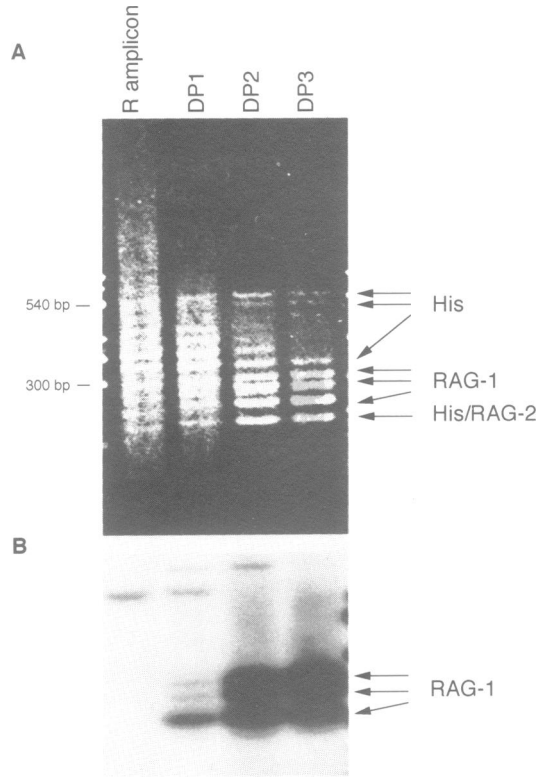

Figure 2. **(A)** Difference products obtained from the subtraction of the control line 3TGHR from the transfectant TXH8. TXH8 stably expresses RAG-1 and RAG-2 and hence V(D)J recombinase activity, while 3TGHR does not express these genes (see text). mRNA was obtained from 3TGHR and TXH8 and subjected to cDNA RDA under standard stringency. Difference products were run on a 1.2% NuSieve agarose/1% agarose gel. Representative amplicon (R amplicon), first, second and third difference products (DP1, DP2 and DP3) are shown. Arrows indicate the identity of the ethidium-stained bands. Products were identified by hybridizing Southern blots with probes for RAG-1 and RAG-2, or histidinol dehydrogenase (His), a transfection marker gene only present in TXH8 [shown for RAG-1 in (**B**)]. The RAG-2 band co-migrates with a His difference product. All bands observed are of the predicted size based on the known sequences of the RAG and His cDNAs. In this and other figures, scanned images of gels and autoradiographs, or computer generated images from a phosphoimager are presented. No processing has been performed, other than to adjust sizes to comply with space constraints.

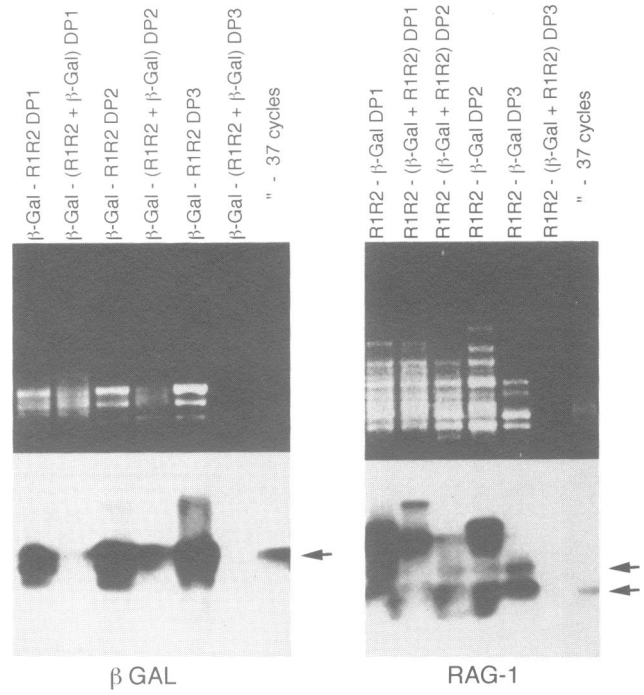

Figure 3. Competitive elimination of abundant difference products by plasmid supplementation. First, second and third difference products (DP1, DP2 and DP3) from βgal minus RAG-1/RAG-2, and RAG-1/RAG-2 minus βgal subtractions, with and without plasmid supplementation. Plasmid supplements (pMFG-βgal, pMFG-RAG-1 and pMFG-RAG-2 digested with Sau3A), are indicated after the driver in parenthesis. The upper panel shows ethidium stained agarose gels from each direction of subtraction. Competed samples show a depletion of specific product to undetectable levels by DP3. The lower panel shows Southern blots of the gels pictured above. βgal minus RAG-1/RAG-2 was hybridized with a probe prepared from the translated region of β-galactosidase. RAG-1/RAG-2 minus βgal was hybridized with a similar probe for RAG-1. Enrichment of amplifiable differences is seen in the uncompeted lanes of both experiments (arrows). The size of fragments capable of efficient amplification typically decreases between DP1 and DP3. Products not hybridizing to RAG-1 in DP3 originate either from the untranslated portion of the RAG-1 message, or are specific to RAG-2 (not shown).

lines, and northern and Southern blots performed. Probing these with RAG-1, RAG-2 or βgal confirmed that all three genes were expressed at an extremely high level (data not shown). The efficiency of infection was monitored by staining MFG-βgal infected cells with Xgal and found to be >95%. In this experiment, the double infectant MFG-RAG-1/MFG-RAG-2 was used as the tester, and MFG-βgal as the driver, to ensure that viral-related messages were not cloned.

The competing plasmids (5 μg of either pMFG-RAG-1 and pMFG-RAG-2 or pMFG-βgal) were cut with Sau3A to generate identical restriction fragments to those found in the tester and were introduced into the tester:driver mixture at the hybridization step. RDA was otherwise performed as normal (Figure 3). The ethidium stained gels show that in contrast to the usual stepwise enrichment of difference products evident in the uncompeted lanes, competed samples show a relative depletion of specific product.The higher ratio of competitor to product in the hybridization leading to DP3 is sufficient to prevent detection of these species in DP3 after the usual 32 cycles of amplification (Figure 3). Some breakthrough is detectable by hybridization after an extra five cycles (37 cycle lanes). Despite their high initial levels, RAG-1, RAG-2 and βgal messages were practically eliminated by the third difference product, although no new RDA products were isolated as a result. Plasmid competition therefore offers a practical method by which a limited number of messages may be eliminated from a difference product.

Detection of up-regulated messages by cDNA RDA

The protocol for genomic RDA requires that a vast excess of driver is used in order to compete out all sequences present in both representations. In cDNA RDA, however, it would be greatly advantageous to be able to detect differentially regulated genes, which would be present, albeit at different levels, in both representations.

Originally, we reasoned that in order to achieve this it would be necessary to deplete the tester and driver amplicons of low abundance sequences. Fragments present in low abundance in the driver should be eliminated, while the same fragments present at up-regulated levels in the tester should remain, permitting

detection by subtraction. This was accomplished by a process of melt-depletion of linker-ligated cDNA in which unmixed samples were melted at 98°C and allowed to reanneal for 1 h, a duration determined empirically to allow much more efficient reannealing for more abundant species than for rarer species (12). Amplification of the annealed product by PCR generated a population with considerable bias against low abundance sequences, as only annealed fragments were capable of amplification (data not shown). This effectively amplifies the differences to make the depleted representations better substrates for RDA.

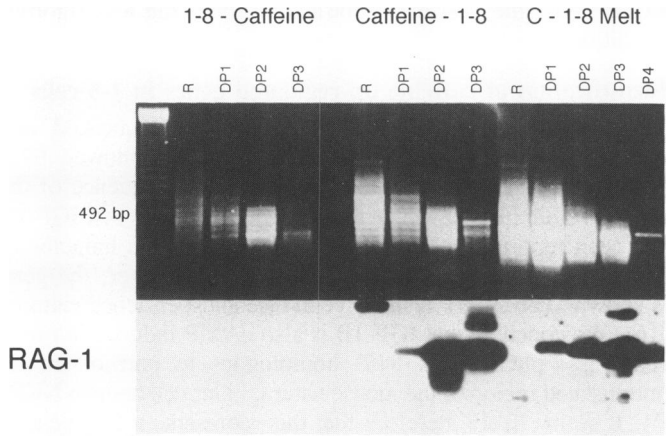

Figure 4. Enrichment of caffeine up-regulated RAG-1 by cDNA RDA. Poly A^+ mRNA was prepared from 1-8 pre-B cells stimulated for 12 h with 1 mM caffeine or from unstimulated control cells, and reverse transcribed. The resulting cDNA was analyzed by low stringency cDNA RDA. Subtractions were performed in both directions (Caffeine−18; 18−Caffeine) and with melt depletion (C−18 Melt). See text for a description of melt depletion. Southern blots of representations (R) and of difference products DP1−DP3 were probed with RAG-1. Upper panels show ethidium stained agarose gels (1.3%). Lower panels show the corresponding Southern blots probed with RAG-1. Enrichment was obtained, with or without melt depletion, in the Caffeine−18 direction but not in the opposite (control) direction.

Melt-depletion alone would be insufficient to overcome the extremely high driver to tester ratios employed in the second and third rounds of hybridization. Consequently, the tester:driver ratio for these hybridizations was increased to 1:100, 1:400, 1:80 000, 1:800 000 for DP1, DP2, DP3 and DP4, respectively, compared to the normal ratios for genomic RDA of 1:100, 1:800, 1:400 000, 1:8 000 000. As a test system we used the cell line 1-8 (8), in which RAG-1 and RAG-2 messages are known to be up-regulated by caffeine (13).

Double stranded cDNA was prepared from two populations of 1-8 cells, one of which had been exposed to 1 mM caffeine for 12 h. RDA was performed, with the above modifications, in forward and reverse subtractions (i.e. induced minus uninduced and vice versa) from both melt-depleted and undepleted representations. As expected, a greater degree of heterogeneity of difference products was observed under the less stringent conditions. However, clear bands were visible in third and fourth difference products. Encouragingly, these differed significantly between directions of subtraction (Figure 4). Southern blots of the difference products were probed with RAG-1 to assess whether up-regulated products were being preferentially amplified. cDNA RDA resulted in the amplification of RAG-1 in the induced minus uninduced direction only, confirming our

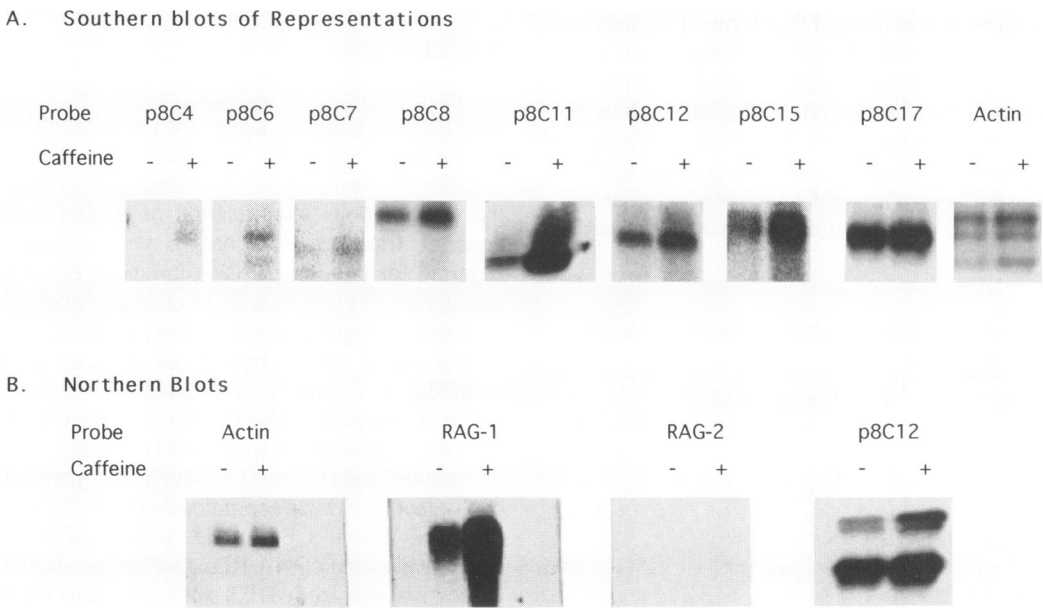

Figure 5. Clones obtained by low stringency cDNA RDA are differentially present in driver and tester amplicons. The fourth difference product of the Caffeine minus 1-8 subtraction was cloned into pBluescript KS^+II, and clones picked at random for analysis. (**A**) Southern blots of representations probed with inserts from cloned difference products. Six out of eight clones, p8C4, p8C6, p8C7, p8C11, p8C12 and p8C15, proved to be caffeine up-regulated. Exposure times vary. (**B**) The validity of this approach was confirmed by probing against northern blots containing 2 μg of $polyA^+$ mRNA per lane (caffeine induced and uninduced). Actin, RAG-1, RAG-2 and p8C12 probings are shown. RAG-1 and RAG-2 were clearly up-regulated by caffeine, with no alternative splicing noted. This was also the case for p8C11 and p8C15 (not shown). p8C4, p8C6 were only barely detectable by northern hybridization. p8C12 and p8C7 exhibited upregulation of only the the upper of two presumably differentially spliced messages.

predictions (Figure 4). Somewhat to our surprise, RAG-1 was obtained regardless of whether the original representations had been melt depleted, indicating that, at least in this case, higher tester:driver ratios in the hybridizations were sufficient to permit selective enrichment of the up-regulated product. A repetition of the experiment produced the same results, indicating the reproducible nature of the technique.

To confirm that our observations were the result of less stringent hybridization, we performed the same subtraction using standard cDNA RDA conditions. This resulted in a DP3 which lacked any identifiable difference products. It was necessary to load the entire DP3 (a hundred times the normal loading) to detect a background smear of non-specific products which was identical for both directions of subtraction (data not shown). This implies that there were no genes detectable by this technique which were 'absolutely' controled by caffeine, and that up-regulated products could only be detected by the more relaxed conditions.

There were clearly many other species present in the third and fourth difference products of the induced minus uninduced subtractions. To assess whether these were derived from caffeine sensitive mRNAs, a number of fragments were cloned at random and prepared as probes against a Southern blot of starting representations. In the induced minus uninduced direction, six out of eight clones were found to be caffeine up-regulated by 2-fold or more (Figure 5A). These probes gave similar patterns on a northern blot of the original RNA, indicating that representation blots accurately reflected the starting levels of mRNA (Figure 5B and data not shown). Interestingly, the levels of cellular mRNA for the species detected varied considerably, from abundant (p8C12) to barely detectable (p8C4), suggesting that detection is not limited to a small proportion of mRNA species. Similarly, the degree to which each species was up-regulated was highly variable, ranging from 2- to 80-fold (Table 1). On the northern blots, it was noticed that some of the species detected were present as differentially spliced forms which were affected unequally by caffeine treatment (e.g. p8C12, Figure 5B). This may partly account for the detection of species showing relatively low levels of up-regulation. However, differential splicing is certainly not a prerequisite of detection, as RAG-1 was clearly up-regulated approximately 10-fold by caffeine and was enriched by cDNA RDA. This indicates that it is possible to preferentially detect messages up-regulated from basal levels of expression by 5–10-fold. Detection of differentially spliced forms is also a useful feature of the method. The non-up-regulated clones p8C8 and p8C17 were both highly abundant species (p8C8 is a component of a mouse Na/K ATPase subunit), which may explain why they escaped subtraction under the less rigorous conditions.

Table 1. Degree of up-regulation of messages detected by clones isolated by cDNA RDA from caffeine-induced 1-8 cells

Clone (induced–uninduced)	Fold up-regulated*
p8C4	80
p8C6	12
p8C7	2.5
p8C8	1.5
p8C11	5
p8C12	2
p8C15	2
p8C17	1.1

Representation amplicons were blotted and probed with the inserts of the isolated difference products listed. The hybridized filter was exposed in a phosphoimager cassette (Molecular Dynamics), and the resulting signal quantified in the linear range using the ImageQuant software package. Values are normalized against mouse β-actin message (which revealed slightly higher loading of the induced lane; see Figure 6). *Uninduced message level is 1. Data were corroborated by northern blots of the original mRNA samples, probed with the same difference products.

Identification of caffeine up-regulated genes in 1-8 cells

The six caffeine up-regulated clones were sequenced and compared with the GenBank database. p8C11 showed 99% homology with bases 295–558 in the coding sequence of the mouse insulin like growth factor 1B gene (IGF-1B) (14). IGF-1B has been reported in B cells, although its caffeine inducibility in these cells has not been recorded (15). However, the gene is known to be cAMP responsive in osteoblast enriched cultures (16). We conclude that IGF-1B is also cAMP induced in pre-B cell lines. p8C15 was 84% homologous to part of the 3′ untranslated region of the mouse natural killer cell antigen NKR-P1. It seems likely therefore that this represents a fragment of a novel member of the NKR-P1 gene family. If so this would be the first demonstration of such a message in pre-B cell lines. Probing of a northern blot of RNA from a variety of cell lines revealed high levels in uninduced fibroblasts, but little or no message in uninduced pre-B cell lines. However, the gene is induced by retinoic acid in an embryonal carcinoma cell line (P19), tentatively suggesting a role in differentiation events (data not shown). p8C4, p8C6, p8C7 and p8C12 were fragments of novel genes, exhibiting no significant homology to any known sequence. Examination of Northern blots, and their different extents of up-regulation indicates that they are the products of distinct genes.

DISCUSSION

We have extended the application of the PCR-coupled subtractive hybridization process, RDA, for use with cDNA and have made a number of important modifications and observations which enhance the usefulness of the technique. cDNA RDA facilitated the unambiguous identification of RAG-1 and RAG-2 from a genomic transfectant in only 3 weeks, compared to 2 years required for the original isolation using conventional molecular approaches (11). cDNA RDA was also applied to the detection of genes up-regulated by caffeine in a pre-B cell line, producing results which would have been difficult to obtain by other methods.

The technique has also proved to be very sensitive, as revealed by an experiment with CB17 SCID (severe combined immunodeficient) mice (17) which lack mature lymphocytes due to a defect in V(D)J recombination (18). Tester and driver were prepared from CB17 and CB17 SCID kidney cDNA, respectively, and cDNA RDA performed. Half of the 40 random clones isolated from DP3 were shown to be derived from the variable portion of functionally rearranged immunoglobulin kappa transcripts, the only possible source of which was lymphocytes transiting the kidney at the time of isolation (data not shown). This demonstrates that cDNA RDA is capable of isolating genes expressed in only a very small fraction of cells (substantially less than 1%) from which the tester is derived.

Although the alterations to the basic RDA procedure were relatively minor, they have fundamental operational implications.

RDA was developed primarily for identification of genomic differences, which, when combined with a careful breeding program or a well known pedigree, has resulted in the identification of markers linked to a number of genetic traits, including *nude* and *staggerer* (3,19). At the genomic level, this requires a reduction in complexity, which is created by amplifying DNA digested with restriction enzymes having six base pair recognition sequences to generate limited representations. Differences are often detected because the homologue of an amplifiable target in the tester is present on a larger, unamplifiable, restriction fragment in the driver. In cDNA RDA, where the complexity is far lower, the representation can be made much more complete by the use of a four-cutter restriction enzyme. Under these conditions, a larger fragment may still be amplifiable, and therefore be present in the driver to subtract out the target. Therefore the basis for subtraction in cDNA RDA is predominantly the presence or absence of message, rather than differences in restriction fragment size. The increased representation is, however, highly beneficial, since one four-cutting digest covers the vast majority of mRNA species present in the initial population.

In addition to the increased representation, there are a number of other advantages in substituting cDNA for genomic DNA as the starting material for difference analysis. First, the gene or genes responsible for any phenotype which results from lack of expression may be directly detected by cDNA RDA. Genomic RDA may, for example, detect many fragments of a large deletion, whereas cDNA RDA detects only missing expressed sequences, which can be rapidly cloned and used to isolate full length cDNA for complementation assays. Second, using cDNA RDA it is possible to follow events in a time-dependent manner, identifying new genes as they become expressed. This can be used to compare developmental stages, cell cycle phases, or to follow the progression of expression proceeding a particular event, for example a stimulation of cells with a growth factor, or an insult to cells by heat shock. Third, in identifying the genes responsible for a particular phenotype, cDNA RDA can detect both the genes, and any downstream products dependent upon the altered regulation. In this way, it should be possible to build up a more complete picture of the event from one initial set of data.

Differential display PCR (DDPCR) (2) is the most comparable method of studying expressed genes currently available, and has been employed with some degree of success (20–23). cDNA RDA shares most of the major attractions of DDPCR, and has in addition a number of significant advantages. Like DDPCR (22), cDNA RDA permits the investigation of gene expression during limited time windows, requiring relatively low amounts of starting material. The major advantages of cDNA RDA relate to the kinetic enrichment achieved by the subtractive process. The production of unambiguous difference products eliminates the need for difficult analysis of numerous, mostly uninteresting, products separated by parallel denaturing gel runs using ^{35}S-labeled ATP, which is necessary with DDPCR. The easily cloned, ethidium visible difference products of cDNA RDA are readily confirmed as true differences by probing against the initial representations or against mRNA on northern blots, and we have not experienced significant problems of 'false positives'. Rapid identification by automatic sequencing is also straightforward.

The process of cDNA RDA employs specific, full length (24mer) primers throughout, avoiding the problem of mispriming associated with the permissive conditions required for very short 5′ primers used in DDPCR. In addition, the use of linkers ligated to full length cDNA digests means that a single set of primers are sufficient to generate amplicons representing the majority of cDNA species. In DDPCR, only approximately one in twelve messages are represented by a particular combination of primers (2), and some 20–25 5′ primers combined with four sets of degenerate 3′ primers are required to cover the whole cDNA population (24,25).

By lowering of the stringency of subtraction, cDNA RDA has been used to detect as little as a 2-fold up-regulation from basal expression in a variety of messages. Therefore the technique is currently capable of isolating genes (e.g. RAG-1) which are present at different levels in driver and tester amplicons. The capacity to raise the tester:driver ratio in the hybridizations without necessarily sacrificing the specificity of difference isolation is probably due to advantageous kinetic enrichment after the first subtractive hybridization. Future refinements should increase this enrichment by competing out persisting high abundance messages by driver supplementation.

Finally, cDNA RDA is not, in theory, dependent on the isolation of polyA$^+$ mRNA. This would permit its use to identify differences in non-polyadenylated mRNA species, and we are currently examining the possibility of applying cDNA RDA to prokaryotic systems. A number of other refinements are also being explored in our laboratory, including the elimination of large numbers of known messages (for example from a partially characterized deletion mutant) by supplementing the driver with full length cDNAs. These would be generated by RT-PCR (or by using a cloned cDNA if available), and added at a concentration comparable to that found in the tester (determined by Northern blotting), prior to creation of the representation.

Like any technique, cDNA RDA is not comprehensive, and it is important to appreciate some of its limitations to avoid inappropriate application. The process is unlikely to identify differences due to point mutations, very small deletions or insertions, fragments from the ends of transcripts, or fragments which lack appropriate enzyme sites. Evidence from the competition experiment, for example, is compatible with the theory that the RAGs do not activate the transcription of new genes, as no difference products were found when RAG sequences were competed away (Figure 3). However, due to the above limitations, it is not possible to draw negative conclusions from cDNA RDA. Most of these limitations apply equally to any other available technique.

Alternative gene expression is responsible for phenotypes which differ between otherwise similar populations. Identification of the genes responsible is facilitated by rapid reduction of the numbers of expressed genes to a few candidates which differ between the isolates. We have described a rapid and effective method by which this can be achieved. Both here, and in ongoing research, we have found the technique reliable and sensitive. In addition to the work described, cDNA RDA has been successfully applied to a number of biologically interesting systems, including the isolation of candidate genes necessary for efficient V(D)J recombination (manuscript in preparation). Together, cDNA RDA and differential display PCR will make major contributions to the characterization of metabolic events at the molecular level.

ACKNOWLEDGEMENTS

The authors thank N.Lisitsyn for advice and a detailed protocol for genomic RDA, M.Sadelain, D.Silver, D.Baltimore and

R.Mulligan for retroviral producer cells and plasmids, and members of the laboratory for helpful discussions. M.H. and D.G.S. were supported by the Howard Hughes Medical Institute.

REFERENCES

1. Wieland, I., Bolge, G., Asouline, G. & Wigler, M. (1990) Proc. Natl. Acad. Sci. U.S.A. **87**, 2720–2724.
2. Liang, P. & Pardee, A. B. (1992) *Science* **257**, 967–971.
3. Lisitsyn, N., Lisitsyn, N. & Wigler, M. (1993) *Science* **259**, 946–951.
4. Bishop, D. T., Williamson, J. A. & Skolnick, M. H. (1983) *Am. J. Hum. Genet.* **35**, 795–815.
5. Alberts, B., et al. (1983) Molecular biology of the cell (Garland Publishing, Inc., New York).
6. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) Molecular Cloning, a laboratory manual. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor).
7. Ausubel, F. M., *et al.* (1989) Current Protocols in Molecular Biology (Greene Publishing Associates and Wiley-Interscience, New York).
8. McKearn, J. P. & Rosenberg, N. (1985) *Eur J Immunol* **15**, 295–298.
9. Schatz, D. G. & Baltimore, D. (1988) *Cell* **53**, 107–115.
10. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) *J. Mol. Biol.* **215**, 403–410.
11. Schatz, D. G., Oettinger, M. A. & Baltimore, D. (1989) *Cell* **59**, 1035–1048.
12. Britten, R. J. (1966) *Science* **154**, 791–794.
13. Menetski, J. & Gellert, M. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 9324–9328.
14. Bell, G. I., Stempien, M. M., Fong, N. M. & Rall, L. B. (1986) *Nucleic Acids Res.* **14**, 7873–7882.
15. Baier, T. G., Jenne, E. W., Blum, W., Schonberg, D. & Hartmann, K. K. (1992) *Leukemia Research* **16**, 807–814.
16. McCarthy, T. L., Centrella, M. & Canalis, E. (1990) *J Biol Chem* **265**, 15353–15356.
17. Bosma, G. C., Custer, R. P. & Bosma, M. J. (1983) *Nature* **301**, 527–530.
18. Schuler, W. & Bosma, M. J. (1989) *Curr. Top. Microbiol. Immunol.* **152**, 55–62.
19. Lisitsyn, N. A., *et al.* (1994) *Nature Genetics* **6**, 57–63.
20. Donohue, P., Alberts, G. F., Hampton, B. S. & Winkles, J. A. (1994) *J. Biol. Chem.* **269**, 8604–8609.
21. Jung, M., Kondratyev, A. D. & Dritschilo, A. (1994) *Cancer Research* **54**, 2541–2543.
22. Nishio, Y., Aiello, L. P. & King, G. L. (1994) *FASEB Journal.* **8**, 103–106.
23. Sager, R., Anisowicz, A.M.N., Liang, P. & Sotiropoulou, G. (1993) *FASEB Journal.* **7**, 964–970.
24. Bauer, D., *et al.* (1993) *Nucleic Acids Res.* **21**, 4272–4280.
25. Liang, P., Averboukh, L. & Pardee, A.B. (1993) *Nucleic Acids Res.* **21**, 3269–3275.