Nucleic Acids Research, 1998, Vol. 26, No. 12 **3059–3065**

Detection of differentially expressed genes in primary tumor tissues using representational differences analysis coupled to microarray hybridization

Scott M. Welford[1], Jeff Gregg[2], Emily Chen[2], Darryl Garrison[3], Poul H. Sorensen[5], Christopher T. Denny[1,3] and Stanley F. Nelson[1,3,4,*]

[1]Molecular Biology Institute and the Departments of [2]Pathology, [3]Pediatrics and [4]Biological Chemistry, UCLA Medical Center, Los Angeles, CA 90095, USA and [5]Department of Pathology, British Columbia Children's Hospital, Vancouver, BC, Canada

Received December 4, 1997; Revised and Accepted April 28, 1998

ABSTRACT

The identification of differential gene expression between cells is a frequent goal in modern biological research. Here we demonstrate the coupling of representational difference analysis (RDA) of cDNA with microarray analysis of the output for high throughput screening. Two primary Ewing's sarcoma tissue samples with different biological behavior *in vivo* were compared by RDA: one which was metastatic and progressed rapidly; the other localized and successfully treated. A modified RDA protocol that minimizes the necessary starting material was employed. After a reduced number of subtractive rounds, the output of RDA was shotgun cloned into a plasmid vector. Inserts from individual colonies from the subtracted library were amplified with vector-specific primers and arrayed at high density on glass slides. The arrays were then hybridized with differentially fluorescently labeled starting amplicons from the two tissues and fluorescent signals were measured at each DNA spot. We show that the relative amounts of fluorescent signal correlate well with the abundance of fragments in the RDA amplicon and in the starting mRNA. In our system, we analyzed 192 products and 173 (90%) were appropriately detected as being >2-fold differentially expressed. Fifty unique, differentially expressed clones were identified. Therefore, the use of RDA essentially provides an enriched library of differentially expressed genes, while analysis of this library with microarrays allows rapid and reproducible screening of thousands of DNA molecules simultaneously. The coupling of these two techniques in this system resulted in a large pool of differentially expressed genes.

INTRODUCTION

The ability to determine mRNA expression differences between cell lines and tissues is a powerful tool in modern biological research. Several recent and rapid PCR-based methods, including subtractive suppressive hybridization (SSH), representational differences analysis (RDA) and differential display (1–4), have been proposed for the cloning of genes which are differentially expressed between two tissues. These methods have been validated and successfully applied to various research problems. Recently, cDNA microarrays and oligonucleotide arrays have been developed and used to quantitate differential gene expression by hybridizing a complex mRNA-derived probe onto an array of PCR products or oligonucleotides representing specific cDNAs (5–7). Microarrays allow thousands of genes to be monitored simultaneously for expression level and compared between many different tissues. However, even arrays of thousands of known genes are likely to miss key differentially expressed genes, as only a fraction of all potentially expressed genes can be currently arrayed on glass slides. Furthermore, oligonucleotide arrays rely entirely on primary EST/gene sequencing data and are not widely available at reasonable cost. To date, probably <50% of all human cDNA sequences are present in the public EST database (Unigene). Thus, there is still considerable room for cloning of novel genes involved in a myriad of cellular pathways. Therefore, differential cloning techniques that give the ability to clone specific genes that are meaningfully differentially expressed and can be rapidly and exhaustively screened are desirable. This allows maximal focus on the tissue of interest and includes genes which are not yet in the public database.

RDA has been successfully adapted to identify genes that are differentially expressed between two populations of cells (4) and has been successfully used in cell line experiments (8–14). Representative cDNA fragments from each population are first generated by restriction endonuclease digestion of cDNAs

*To whom correspondence should be addressed. Tel: +1 310 794 7981; Fax: +1 310 206 9819; Email: snelson@ucla.edu

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors

followed by PCR amplification. The resulting mixtures, termed 'amplicons', are then subject to successive rounds of subtractive cross-hybridization followed by differential PCR amplification. This leads to progressive enrichment of cDNA fragments that are more abundant in one population than the other.

While initial experience has been encouraging, a number of critical determinants have emerged from our experience with RDA. First, the relative difference in expression level does impact on whether a gene is recovered by RDA. In a mixed population of differentially expressed genes in which some genes differ by 5-fold and others by 100-fold, the 100-fold genes will be amplified preferentially in RDA. However, absolute expression level has not been a crucial determinant for identifying genes, because fragments from both abundant and rare transcripts have been isolated (15). Second, like all subtractive methodologies, RDA can only compare cell populations two at a time. Third, each RDA recovers typically only 6–12 differentially expressed cDNA fragments, which likely represent only a small subset of all the differentially expressed genes (4,8–15). Finally, RDA has generally required large (~100 µg) quantities of starting RNA, which precludes using non-renewable cell sources such as tumor specimens.

Here we incorporate alterations in the RDA protocol that allow smaller starting tissue amounts, which allows its application to biopsy or otherwise small specimens. RDA is halted at an early round of subtraction at which there has been both enrichment for differentially expressed sequences and preservation of a wide distribution of the differentially expressed genes. This more complex output of RDA is shotgun subcloned into a plasmid vector and the inserts are microarrayed onto dozens of glass slides at discrete positions. The arrayed inserts are subsequently co-hybridized with differentially labeled amplicons to identify those inserts which are differentially represented in the starting populations. After the analysis of only 192 products, 50 differentially expressed genes were found. Therefore, the merging of RDA with microarray analysis proved to be an efficient method for detection of unique, differentially expressed genes.

MATERIALS AND METHODS

Tissues

Serial sections of two fresh frozen Ewing's sarcoma tumor specimens from separate patients were used, termed ES7 and ES10. Tissue culture cell lines consisted of the NIH 3T3 cell line transfected with either an EWS/FLI1 transforming variant (clone 10.10) or a plasmid control (clone 46.1) (16). Cell lines were cultured in DMEM with 5% calf serum and glutamine (Gibco BRL).

Isolation of total RNA and cDNA synthesis

Total RNA was extracted from 100 mg fresh frozen tumor tissue with Stat-60 reagent using the manufacturer's recommendations (Tel-Test Inc.). cDNA was synthesized from 30 µg total RNA by oligo(dT) priming using Superscript II as recommended by the manufacturer (Gibco BRL).

Generation of the subtracted library by RDA

RDA procedures were similar to previously published methods (15), with some notable exceptions. Poly(A)$^+$ RNA was purified from only 30 µg total RNA using the Oligo-tex mRNA Purification Kit (Qiagen). RDA amplicons made from *Dpn*II-digested cDNA were generated by PCR and purified on silica spin columns (Qiaquick PCR Purification Kit; Qiagen). Subtractive hybridizations were performed in reduced volumes (2.5–5 µl) using 10% PEG (final concentration) and less amplicons (5–10 µg). Tester to driver ratios were preserved and, thus, decreased amounts of tester were used. The mung bean nuclease digestion used during differential amplification steps in previous protocols was omitted. In the Ewing's sarcoma RDA, both ES7 – ES10 and ES10 – ES7 subtractions were performed. The ratios for rounds 1 and 2 were 1:100 and 1:500 respectively. Subtracted RDA products were digested with *Dpn*II and shotgun cloned into pBluescript KS II+ (Stratagene).

Analysis of RDA subtraction by SAGE

We assessed the diversity of RDA output at earlier stages of subtraction (i.e. following two rounds) by serial analysis of gene expression (SAGE) (10). Adapters containing a type II restriction enzyme recognition site to *Bsm*FI (GGGAC/CCCTG) and an anchoring enzyme recognition site to *Dpn*II (GATC/CTAG) were ligated to RDA fragments. The SAGE-specific adapters used to create the libraries were: 5′-TTTGGATTTGCTGGTGCAGTA-CAACTAGGCTTAATAGGGAC-3′ and 5′-GATCGTCCCTAT-TAAGCCTAGTTGTACTGCACCAGCAAATCC*-3′; 5′-TTTC-TGCTCGAATTCAAGCTTCTAACGATGTACGGGGAC-3′ and 5′-GATCGTCCCCGTACATCGTTAGAAGCTTGAATTCGAG-CAG*-3′ (* = 3′ amino modifier C7). Digestion with *Bsm*FI liberated a 14 bp tag from the ends of each RDA fragment. These tags were then concatamerized, amplified, subcloned into pBluescript KS II+ and sequenced.

Microarraying of cloned RDA products

After shotgun cloning the RDA products into pBluescript KS II+, individual colonies from the subtracted libraries were picked into 96-well liquid cultures. Plasmid inserts were PCR amplified using amino-modified vector-specific primers (5′-GGCCGCTC-TAGAACTAGTGGAT-3′ and 5′-CTCGAGGTCGACGGTAT-CGATA-3′) which amplify the inserted fragment from pBluescript KS II+. PCR was performed by adding 0.5 µl saturated growth liquid culture to 50 µl PCR reactions containing 10 mM Tris (pH 9.0), 50 mM KCl, 0.1% gelatin, 2.5 U Taq DNA polymerase and 150 µM dNTP in 96-well plates (MJ Research). Thermal cycling conditions consisted of an initial denaturation at 94°C for 2 min, followed by 35 cycles of 94°C for 1 min, 68°C for 1 min and 72°C for 1.5 min, with a final 10 min 72°C extension, in a PTC100 thermal cycler (MJ Research). Five microliters of each PCR amplification product were examined by agarose gel electrophoresis and ethidium bromide staining. A single band was detected in 187 of the 192 PCR reactions performed (data not shown). Each PCR product was recovered from the remaining 45 µl (~1–2 µg) by ethanol precipitation in the 96-well plates. The PCR products were arrayed onto glass slides in a manner similar to that already published (5). Briefly, the PCR products were resuspended in 15 µl 1× SSC. A custom built arraying robot picked up ~600 nl DNA solution and deposited 1–4 nl DNA solution in duplicate onto a silanized glass slide surface (Sigma). The design of the arrayer is publicly available at (http://cmgm.stanford.edu/pbrown). Up to 84 slides were created per printing. Each slide was hydrated for 10 s over a 37°C water bath, snap dried for 2 s on a 100°C block, then crosslinked with

4000 mJ short wave UV irradiation (Stratalinker; Stratagene). Each slide was then washed for 2 min sequentially in 0.2% SDS and distilled water. The crosslinked slides were denatured in 100°C water for 2 min, desiccated in 95% ethanol and air dried. No specific blocking step was required.

RDA amplicons (2 μg) were fluorescently labeled with either Cy3 or Cy5 dyes. Maximum incorporation of fluorescent nucleotide was accomplished with 100 μM dGTP, dTTP and dATP and 10 μM unlabeled dCTP: dCTP-Cy3 or dCTP-Cy5 (Amersham) was doped in at a final concentration of 20 μM and incorporated into DNA using random primer labeling as per the manufacturer's recommendations (Stratagene). Labeled RDA probe was purified through Sephacryl 200 spin columns (Pharmacia) and precipitated. The labeled probe was resuspended in 20 μl 3× SSC, 1% SDS, 5× Denhardt's solution, 100 mg/ml sheared salmon sperm DNA, 50% formamide and 10% dextran sulfate. The probe was denatured in the hybridization solution at 80°C for 10 min and applied to the arrayed/denatured slide at 42°C for 12–20 h in a humidified chamber. Hybridized slides were washed in 2× SSC, 0.1% SDS for 5 min at room temperature, then 0.2× SSC for 5 min prior to scanning.

Slides were scanned in a custom built two color laser scanning fluorimeter and the image files analyzed with NIHimage v.1.60 to quantitate the signal at each spot. A separate 1024×1024 image was captured for each of the two fluorophores used. The resolution of the scan was 20 μm per pixel with a 256 gray scale.

Northern analysis

Five micrograms of total RNA obtained from ES7 and ES10 tumor sections were electrophoresed in denaturing formaldehyde agarose gels. Following overnight capillary transfer to MSI Nitroplus membranes, the blots were probed with PCR-amplified fragments which were radioactively labeled with the Random Primed DNA Labeling Kit (Boehringer-Mannheim). Hybridization and high stringency washing with a final wash of 0.2× SSC, 0.1% SDS at 52°C for 20 min were performed.

Southern analysis

Driver DNA (800 ng) from each direction was fractionated on a 1.2% agarose gel. The DNA was denatured in the gel by submerging in 0.5 M Tris–HCl, pH 7.5, 1.5 M NaOH for 30 min and neutralized in 1.5 M Tris–HCl, pH 7.5, 1.5 M NaCl for 30 min. Transfer was overnight in 4× SSC to Nitroplus membrane (MSI). Probes, hybridization and washing conditions were as for northern analysis.

RESULTS

A reduction in the number of subtractive hybridizations leads to greater RDA diversity

For the most part, RDA has been successful in identifying only a small subset of genes that are differentially expressed between two cell populations. Potential explanations for not detecting the remaining differentially expressed genes include: (i) the diversity in the starting amplicon pools is low; (ii) differentially expressed gene fragments are lost along with non-differentially expressed gene fragments, after multiple rounds of RDA. SAGE was used to assess the diversity of RDA amplicon pools following two rounds of subtraction hybridization/differential amplification. SAGE is a method for profiling the expressed transcripts by randomly sampling cDNAs in different libraries by sequencing short 10–14 bp tags derived from a defined region within each gene (16). In our system we applied this technique to RDA products by modifying the linkers to accommodate the *Dpn*II restriction ends of RDA fragments. Following addition of these special adapters, 14 bp tags were released and the original protocol was followed. The tags were concatamerized, cloned and sequenced.

We assessed the diversity of RDA in a model system comparing NIH 3T3 cells expressing a transformation-competent variant of the EWS/FLI1 fusion gene under the transcriptional control of an inducible metallothioneine promoter (clone 10.10) versus NIH 3T3 cells containing empty vector (46.1) (17). After two rounds of RDA, SAGE mini-libraries were made from both the 10.10 – 46.1 and the reciprocal 46.1 – 10.10 outputs and sequenced.

The SAGE sequence results indicated that the RDA outputs following two rounds of subtraction were quite diverse. Sixty five SAGE concatamer clones were sequenced, identifying 1710 tags which represented 621 distinct sequences: 553 were detected one to two times, 55 were counted three to 10 times and 13 were counted more than 10 times. It is difficult to determine whether a sequence that was only found one or two times represents a differentially expressed gene, however, 546 tags were detected in either one of the RDA outputs but not the other. Thus, it was apparent that after two rounds of RDA strong diversity still remained in the subtracted pool.

Optimization of hybridization conditions allows comparison of primary tissues

Reducing the amount of starting tissue required for RDA is necessary to apply RDA to biopsy specimens. Various refinements were incorporated into the RDA procedure to accomplish this (see Materials and Methods). Most importantly was the use of volume excluding agents in the subtractive hybridization so that the amount of amplicons used per round of RDA was reduced by 8-fold (i.e. from 40 to 5 μg).

Our EWS/FLI1-NIH 3T3 model system was used to confirm that the changes in hybridization conditions and number of RDA cycles described in this manuscript would not impair the ability to identify differentially expressed genes. Using a standard methodology and three cycles of RDA, we had previously identified a cohort of genes up-regulated by the EWS/FLI1 fusion gene in NIH 3T3 cells (15). Amplicons were prepared from polyclonal NIH 3T3 populations expressing the EWS/FLI1 fusion gene or empty vector and subjected to two rounds of RDA with reduced amounts of driver amplicons. Products were then shotgun subcloned into plasmids, picked into 96-well plates and amplified inserts from 1512 clones were gridded onto microarrays. These arrays were then hybridized with probes from three different genes known to be up-regulated by EWS/FLI1. While all three genes are undetectable in NIH 3T3 cells without EWS/FLI1, they reach different absolute levels when induced: (i) EAT1, high (approximately equal to β-actin); (ii) stromelysin-1, intermediate; (iii) EAT2, low (~10-fold less than β-actin). All three genes were detected on EWS/FLI1 – empty vector microarrays. Their frequency on the microarray approximated the relative levels of expression: EAT1 was represented in 268 of 1512 microarray spots (17%); stromelysin -1 in 141 (9%); EAT2 in 41 (2.7%). These data confirm that enrichment of both high level and low level differentially expressed genes is achievable

using our new conditions. However, such a model system cannot gauge the idiosyncratic effect of PCR amplification itself. Given a heterogeneous population of cDNA fragments from differentially expressed genes, some may be amplified more efficiently and therefore be more highly represented on the microarrays.

To demonstrate the feasibility of using small amounts of clinical tissue specimens as starting material in our modified RDA protocol, two Ewing's sarcoma tumor specimens from patients with very different biological behavior and clinical outcomes were compared: ES7, from a patient with localized disease who is alive and well; ES10, from a patient with a quickly metastatic tumor who died. Amplicons were generated from 30 µg total RNA from each tumor that was harvested directly from frozen tissue blocks. Each round of RDA subtraction was analyzed by gel electrophoresis and indicated abundant bands which were unique to the ES7 – ES10 or the ES10 – ES7 subtractions, as well as a diffuse smear of products from 200 to 600 bp (data not shown). The heterogeneous products of the second round of RDA were used as substrates in the subsequent experiments.

Differential hybridization of the amplicons to the microarray predicts differential expression of genes

SAGE data from our NIH 3T3 model system suggested that stopping RDA after two rounds lessens the loss of differentially expressed cDNAs, perhaps due to PCR amplification preferences, and maintains diversity. However, it seemed likely that these RDA populations would also have a higher proportion of non-differentially expressed species than RDAs performed using a greater number of rounds. Gridded microarrays were used as a screening technique to identify those RDA clones that were derived from differentially expressed genes.

After two rounds of RDA, ES7 – ES10 and ES10 – ES7 RDA fragments were endonuclease digested and shotgun subcloned into plasmids. Individual clones were picked into 96-well microtiter dishes and inserts were PCR amplified with vector-specific primers. 192 of the RDA products (96 from each subtraction) were arrayed in duplicate on glass slides and co-hybridized with Cy3-labeled ES7 and Cy5-labeled ES10 amplicons. After stringent washing, hybridization signals from each fluorophore were quantitated in the two color fluorescent scanner. Each image was captured separately and each spot was quantitated relative to the local slide background in arbitrary fluorescent units. The net fluorescent signal at each spot from the Cy5 and Cy3 dyes was compared. Overlaying false colored images from the fluorescent signal on a single slide allowed a quick visual inspection and demonstrated that the majority of RDA products hybridized differentially to either the ES7 or ES10 starting amplicons (Fig. 1). The relative fluorescent signals from each probe were measured to determine the relative abundance of each RDA clone in the original amplicon. The vast majority (173/192) showed greater hybridization to the amplicon used as a tester in the RDA (Table 1).

Northern and Southern experiments were performed on selected RDA clones to correlate microarray hybridization patterns with cDNA fragments from differentially expressed genes. Clones from both ES7 – ES10 and ES10 – ES7 RDAs that gave differential microarray signals were radiolabeled and hybridized to Southern blots containing ES7 and ES10 amplicons as well as northern blots containing starting RNAs (Fig. 2). There were both high and low abundance genes in this set and northern and Southern hybridization patterns were completely concordant with the microarray data.

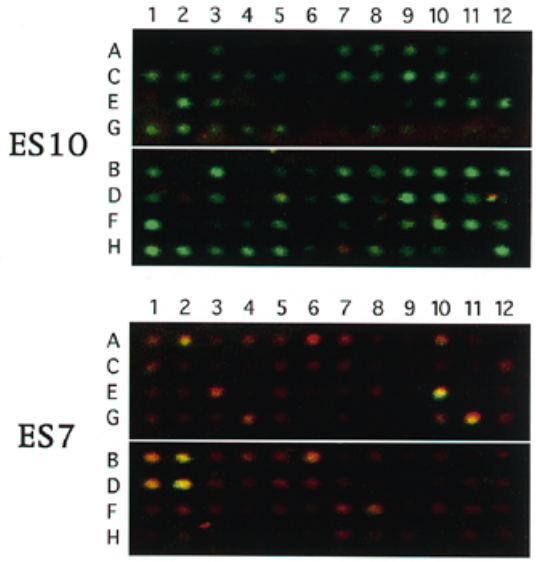

Figure 1. Microarray analysis of 192 RDA products. 192 RDA products were amplified from plasmid clones using vector-specific primers. ES10 indicates the 96 clones derived from the ES10 – ES7 RDA subtraction and ES7 indicates the 96 clones derived from the ES7 – ES10 subtraction. One microgram of ES7 amplicon was labeled with Cy3 and 1 mg ES10 amplicon was labeled with Cy5 by direct incorporation of fluorescently tagged nucleotide. The labeled amplicons were mixed and co-hybridized onto the 192 RDA array. After stringent washing the array was scanned with a laser scanning fluorimeter. Cy3 signal and Cy5 signals were captured separately as a 256 gray scale. The images were overlaid in NIHimage v.1.60 with Cy3 signal false colored red and Cy5 false colored green.

Table 1. Relative hybridization signal of amplicons onto 192 subtracted RDA products

Relative signal	Clones derived from ES10 – ES7	ES7 – ES10
ES10 > ES7	82	0
ES7 > ES10	5	91
ES7 = ES10[a]	4	5
No signal	5	0

[a]Relative signal is <2-fold different for the two probes.

In order to measure how many distinct differentially expressed mRNA were represented, all 192 clones were analyzed by either sequencing or back-hybridization. First, 17 of the most intense signals were sequenced, yielding 10 unique fragments. Using combinations of these fragments as probes on the array (back-hybridization), 99 of the 192 spots were hybridized. Specifically, two of the clones accounted for 37 spots, while eight others hybridized with 62 spots (data not shown). Further sequencing or back-hybridization of the remaining differentially hybridizing spots revealed an additional 40 unique clones. Not surprisingly, the highest degree of redundancy within the clones occurred in those with the highest signal intensity. The more abundant RDA

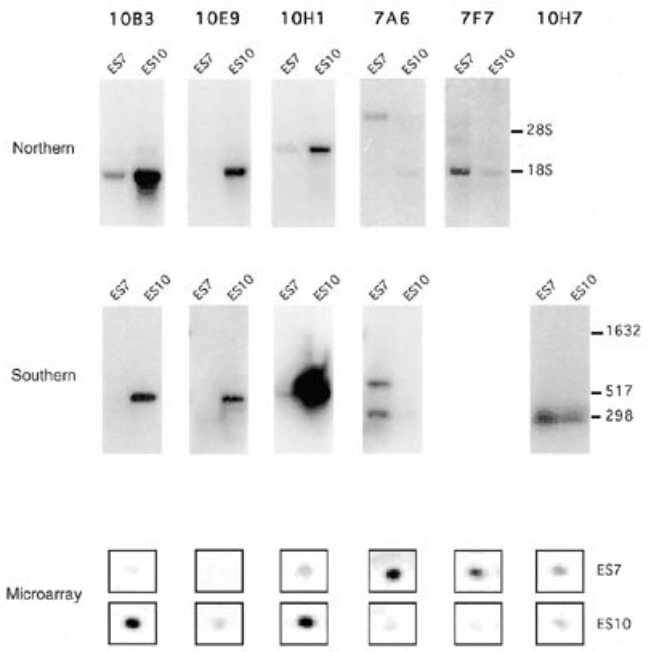

Figure 2. Northern and Southern analyses confirm microarray accuracy. Six clones detected on the microarray as differentially present in either driver or non-differentially present were PCR amplified and used as probes on northern and Southern blots. Fragments were assigned names according to the direction from which they were cloned (10 or 7), followed by their location on the 192 spot microarray. Enlarged microarray signals are aligned below their respective northern and Southern blots. Fragments 10B3, 10E9, 10H1 and 7A6 all show differential detection on the microarray, which is confirmed on Southern blots. As well, hybridization to total RNA from the tumor tissues directly is strongly correlated with the microarray results. Fragment 7F7 is also detected appropriately on the microarray as compared with the northern blot. Fragment 10H7 demonstrated near equal detection on the microarray (a <2-fold difference). This non-differential hybridization was verified on a Southern blot.

products will be arrayed more redundantly on the array, and their high relative concentration in a random priming DNA labeling reaction will result in high intensity signals on hybridization to the microarrays. Thus, there is a theoretical and empirical correlation between high intensity signals and high frequency of redundancy. The argument follows for rare fragments as well. Once the highly redundant clones were identified and removed from the remaining candidates, a higher rate of identification of unique genes was observed. In the subset of spots with medium signals 20 unique sequences were detected out of 54 DNAs arrayed. In the subset of spots with the lowest signals, >90% proved to be of unique fragments (19 of 20).

While 50 of the 192 spots on our initial array proved to be unique, the high rate of detection of unique genes in the medium and low signal subsets suggested that we were not yet saturating our ability to detect unique genes. To confirm this notion, an additional 1152 RDA clones were arrayed to determine if further unique products could be detected. The array was probed with the 50 previously identified gene fragments and signals were detected at 946 of the 1152 DNA spots (data not shown). Thus, there are 206 microarrayed DNAs not represented in the first set which may represent new unique clones. All of these clones belong to the medium signal or low signal groups. The high signal group of clones had been fully detected in screening of the initial 192 microarray.

Based on the data from probing of the 192 RDA fragment array, we can estimate how many of 206 new spots detected on the larger 1152 clone array are unique. Since 99 of 192 (52%) were derived from the 10 high signal clones, we expect these 10 clones to be present on the 1152 array 594 times. Since medium signals were present at 54 of 192 (28%) DNAs of the original array, we expect 324 spots to be from medium signals. However, there is a 2.6-fold redundancy rate on the 192 array. Assuming this same redundancy continues on the new array, only 39% of the clones are new medium signal DNAs. In other words, 126 (39% of the 324 anticipated) should not be represented already within the 21 medium signal clones already detected, while 198 should hybridize with one of the first 21 medium signal clones detected. Of the low signal clones with only a 1.05 redundancy rate, 95% of the low signal clones may be novel. Since 20 of the original 192 belonged to the low signal class, there should be 120 low level signals on the 1152 clone array and 114 (95% of 120) new clones detectable. We estimate from the values above that the ratio between the medium signal clones and the low signal clones should be 126:114. This results in an estimated 109 (53% of 206) being medium signal clones and 97 (47% of 206) being low signal clones. Each medium signal clone will be present at least 2.6 times on the 1152 array and each low signal clone will be present at least 1.05 times on the array. Therefore, we estimate that maximally 43 more unique medium signal clones and 92 more unique low signal clones will be detected as this subsequent array is screened. These are approximations based on the redundancy rate at the initiation of probing the 1152 clone array. The redundancy rate will increase as we identify more clones and, thus, the 135 new clones anticipated is an overestimate. The key point is that a substantial number of unique differentially expressed genes can be identified by this screening approach after the analysis of only 1344 clones.

DISCUSSION

Our primary goals are to rapidly screen small tissue samples in a pairwise fashion for differential gene expression and to analyze the expression pattern of those cloned genes in a wide variety of tissues. The modifications to RDA implemented in our system allow the ready adaptation of RDA to primary tumors. Halting subtraction at the end of two rounds of RDA allows a broader spectrum of differentially expressed genes to be represented and detected within the subtracted RDA fragments. The screening of RDA output by microarray and two color fluorescent hybridization appears to be a powerful means to rapidly generate and screen for genes that are differentially expressed between two tissues or cell lines.

RDA relies on the restriction digestion of cDNA and ligation of adapters to a PCR-amplified subset of all gene sequences. Since each mRNA is 2–4 kb on average, there is a high likelihood of two *Dpn*II sites at appropriate spacing, such that the intervening gene sequence is amplified. However, since not every sequence is equivalently amplified, there will be sequence bias during the multiple rounds of PCR amplification, resulting in the loss of some PCR products. If this selection were to be severe, it would be difficult to detect many differentially expressed genes with RDA. However, this has not been a major problem in our initial screens. Although 10 fragments represented 52% of the

arrayed products, of the first 192 clones analyzed a diverse population of 50 unique differentially expressed fragments was detected and the subsequent analysis of 1152 additional clones by microarray indicates that at least 100 more clones will be detected. The most highly abundant 10 clones likely represent the low number of difference products likely to have been cloned in the typical RDA subtraction.

The subtraction performed in the reduced volume RDA was efficient, with 90% of the arrayed products being more abundant in tester than driver. As it appears that there is a good quantitative correlation between the analysis of abundance of PCR products in the amplicon and the original mRNA, it may be possible to use RDA amplicons as probes onto microarrays fabricated from the RDA products. This has two advantages. First, rare transcripts may be detectable with the amplicon probe but not by a direct mRNA probe. Second, much less starting mRNA (nanograms) may be needed to generate an amplicon from the cDNA than is needed for probing microarrays with mRNA (1–4 μg) directly. Minimization of mRNA needed to create the probe for hybridization to the microarrays extends the utility of the approach.

The comparison of tissues from two different individuals raises the possibility that some of the RDA fragments may be due to primary DNA sequence differences. The vast majority of polymorphisms in the human genome are single nucleotide changes. Most importantly, polymorphisms within the *Dpn*II restriction sites can create sufficient alterations in the abundance of RDA fragments irrespective of differential mRNA expression by altering the size of the *Dpn*II fragment. However, such effects should be minimal. Single nucleotide differences occur within coding sequences in the human genome at a highest rate of ~1 in 1000 nt (18). Thus, the frequency of changing 1 of the 4-nt *Dpn*II recognition sites at either end of a *Dpn*II fragment, thereby creating a restriction fragment polymorphism, would be about once out of every 125 fragments. Therefore, at a maximum <1% of all RDA amplified fragments will be altered due to polymorphism between the individuals.

The comparison of ES7 and ES10 Ewing's tumors was performed with the biological intent of determining genetic markers of metastatic Ewing's tumors. In order to test that the genes isolated in this experiment are actually consistently differentially expressed in metastatic tumors, further analysis will need to be performed. Screening of these candidate genes over a panel of Ewing's tumors will elucidate which genes show linkage to metastasis. This will also answer questions regarding the quality of the tumor samples and contaminating stromal tissue. Before such experiments are performed, however, no conclusions can be made on the biological relevance of the genes isolated here. Of the 50 genes that we have identified, 21 were already known, 20 had only EST matches in the database and the remaining nine had no matches in GenBank. The list of identified genes is available on-line (http://atrip.mednet.ucla.edu/snelson/rdadata).

It has recently been proposed to screen the output of SSH using high density Southern blots of the cloned products (19). This approach is clearly successful and will likely be useful to laboratories without the need to build a DNA microarrayer. However, creating microarrays has several advantages. First, the arrays can be created more quickly. With the robotic arrayer, 84 slides can be arrayed with up to 1600 DNA spots in under 8 h. This large number of replicates is quite useful. Initially, the clones are screened by co-hybridization of differentially labeled drivers to test the efficiency of RDA subtraction. Subsequently, the arrays can be hybridized with batches of the arrayed clones to select a non-redundant set of fragments. We estimate that screening 1600 RDA fragments should allow a thorough, but not complete, detection of differentially expressed sequences. As tissues may express 10 000–20 000 genes and between closely related cell types ~1–10% are differentially expressed, there should be 100–2000 differentially expressed genes (20). In our experience the screening of 1600 RDA products by microarray allows the sampling of ~200–300 unique genes. Most importantly, these replicate microarrays can be probed with multiple other tissues to assess tissue distribution and relation to other biological variables.

Other subtractive techniques, like SSH, have also been screened by microarray hybridization in a similar format. Although not directly compared with the ES7 and ES10 tissues, a similar rate of unique differentially expressed genes are detected by SSH library screenings (L.Goodlick, J.Gregg and S.Nelson, unpublished data).

Two color fluorescent hybridization reveals only the relative abundance of a given gene contained on the microarray within the two probes applied. Absolute quantitation is not determined. We infer from the hybridization intensity that stronger signals indicate higher expression. However, given variations in the amount of DNA coupled to the glass surface, we may under- or overestimate the absolute expression level. Since both Cy3- and Cy5-labeled probes are co-hybridized onto the same spot, the relative signal between the fluorophores strongly correlates with the relative abundance in the probe solutions. Precise relative quantitation comparing many slides could be attained by doping in a standard control hybridization pool of all RDA products labeled with Cy5. Each test sample would be labeled with Cy3, mixed with same amount of control Cy5-labeled probe and co-hybridized onto microarrays. This would allow correlation of the Cy3/Cy5 signal across numerous probes.

The proposed combination approach allows for a thorough analysis of differential gene expression between tissues. As the biological behavior of tissues must be reflected in some manner by differential gene expression, there is great interest in the gene expression profiling of various tissues, for instance to extend biopsy information. Developing focused sets of genes which are likely to be differentially expressed in particular disease states will be beneficial to this application of the microarray technology, as it is currently not feasible to array the entire assortment of genes in higher eukaryotes and screen for all possible differences. Coupling of RDA subtraction with microarray analysis creates a convenient, high throughput means to profile gene expression patterns.

ACKNOWLEDGEMENTS

This work was supported by grant P30 CA16042-22 and a Stop Cancer Research Career Development Award to S.F.N. S.W. is supported by USPHS National Research Service Award GM07185. J.G. is supported by a Medical Training Fellowship

(2T32GM08243). C.D. is supported by a grant from the National Cancer Institute (CA32737).

REFERENCES

1 Diatchenko,L., Lau,Y.F., Campbell,A.P., Chenchik,A., Moqadam,F., Huang,B., Lukyanov,S., Lukyanov,K., Gurskaya,N., Sverdlov,E.D. *et al.* (1996) *Proc. Natl. Acad. Sci. USA*, **93**, 6025–6030.
2 Liang,P. and Pardee,A.B. (1992) *Science*, **257**, 967–971.
3 Lisitsyn,N., Lisitsyn,N. and Wigler,M. (1993) *Science*, **259**, 946–951.
4 Hubank,M. and Schatz,D.G. (1994) *Nucleic Acids Res*., **22**, 5640–5648.
5 Schena,M., Shalon,D., Davis,R.W. and Brown,P.O. (1995) *Science*, **270**, 467–470.
6 Lockhart,D.J., Dong,H., Byrne,M.C., Follettie,M.T., Gallo,M.V., Chee,M.S., Mittmann,M., Wang,C., Kobayashi,M., Horton,H. *et al.* (1996) *Nature Biotechnol*., **14**, 1675–1680.
7 DeRisi,J., Penland,L., Brown,P.O., Bittner,M.L., Meltzer,P.S., Ray,M., Chen,Y., Su,Y.A. and Trent,J.M. (1996) *Nature Genet*., **14**, 457–460.
8 Chu,C.C. and Paul,W.E. (1997) *Proc. Natl. Acad. Sci. USA*, **94**, 2507–2512.
9 Dron,M. and Manuelidis,L. (1996) *J. Neurovirol*., **2**, 240–248.
10 Edman,C.F., Prigent,S.A., Schipper,A. and Feramisco,J.R. (1997) *Biochem. J*., **323**, 113–118.
11 Fu,X. and Kamps,M.P. (1997) *Mol. Cell. Biol*., **17**, 1503–1512.
12 Gress,T.M., Wallrapp,C., Frohme,M., Muller-Pillasch,F., Lacher,U., Friess,H., Buchler,M., Adlerand,G. and Hoheisel,J.D. (1997) *Genes, Chromosomes Cancer*, **19**, 97–103.
13 Lerner,A, Clayton,L.K., Mizoguchi,E., Ghendler,Y., van Ewijk,W., Koyasu,S., Bhan,A.K. and Reinherz,E.L. (1996) *EMBO J*., **15**, 5876–5887.
14 Niwa,H, Harrison,L.C., DeAizpurua,H.J. and Cram,D.S. (1997) *Endocrinology*, **138**, 1419–1426.
15 Braun,B.S, Frieden,R., Lessnick,S.L., May,W.A. and Denny,C.T. (1995) *Mol. Cell. Biol*., **15**, 4623–4630
16 Velculescu,V.E., Zhang,L., Vogelstein,B. and Kinzler,K.W. (1995) *Science*, **270**, 484–487.
17 Thompson,A.D., Braun,B.S., Arvand,A., Stewart,S.D., May,W.A., Chen,E., Korenberg,J. and Denny,C. (1996) *Oncogene*, **13**, 2649–2658.
18 Li,W.H. and Sadler,L.A. (1991) *Genetics*, **129**, 513–523.
19 von Stein,O.D., Thies,W.G. and Hofmann,M. (1997) *Nucleic Acids Res*., **25**, 2598–2602.
20 Zhang,L., Zhou,W., Velculescu,V.E., Kern,S.E., Hruban,R.H., Hamilton,S.R., Vogelstein,B. and Kinzler,K.W. (1997) *Science*, **276**, 1268–1272.