Representative cDNA Libraries and Their Utility in Gene Expression Profiling

BioTechniques 26:542-550 (March 1999)

W.O. Endege, K.E. Steinmann, L.A. Boardman[1], S.N. Thibodeau[1] and R. Schlegel
Chiron Diagnostics, East Walpole, MA; [1]Mayo Clinic and Foundation, Rochester, MN, USA

ABSTRACT

An increasing interest in gene expression profiles in human diseases has led to the use of microdissected tumors and biopsies in gene discovery approaches. Since many of these clinical samples yield extremely small amounts of RNA, reproducible methods are needed to amplify this RNA while maintaining the original message profile. Using the SMARTÔ cDNA Synthesis Method, we show that high-, medium- and low-abundance transcripts can be amplified in a representative fashion and that the resulting cDNA can also be used as a complex probe to confirm gene expression differences identified by other techniques.

INTRODUCTION

Microgram quantities of poly(A)$^+$ RNA are usually required to construct cDNA libraries (5) and to confirm differential gene expression (6,13). Northern blot analysis of differential gene expression typically requires 10 µg of total RNA or 1–4 µg of poly(A)$^+$ RNA. Generation of amplified RNA probes to confirm differential gene expression is reported to require as much as 2 µg of mRNA or 5–40 µg of total RNA (10,11). It is often difficult to acquire sufficient RNA when working with clinical samples such as needle biopsies and microdissected tumors. The use of such tissues is becoming more common in clinical research, and methods to amplify limiting amounts of RNA while maintaining the original message profile are in great demand. Polymerase chain reaction (PCR)-based cDNA amplification methods (2,3) have been described that can be used to circumvent some of the above problems. However, most PCR-based techniques do not generate full-length cDNA libraries, and it is difficult to determine if the original RNA is amplified representatively. Misrepresentation of transcripts due to over-amplification (3) is a serious concern. In this report, we assessed the composition of Switch Mechanism at 5′ end of RNA Transcripts (SMART™) PCR cDNA Libraries (CLONTECH Laboratories, Palo Alto, CA, USA) to determine whether such amplified libraries *(i)* often contain full-length sequences, *(ii)* can retain the original message profile and *(iii)* can be used as tools for evaluating differential gene expression. Conventional reverse transcription (RT) reactions use either the oligo(dT) primer or random hexamers to synthesize cDNA. The SMART cDNA synthesis technology utilizes a combination of two primers in a single reaction. A modified oligo(dT) primer (CDS III/3′ PCR primer) is used to prime the first-strand reaction, while the SMART III oligonucleotide serves as a short, extended template at the 5′ end of the RNA templates. When the reverse transcriptase reaches the 5′ end of the mRNA, the enzyme switches templates and continues replicating to the end of the SMART oligonucleotide. The resulting single-stranded cDNA often contains the full-length transcript and the sequence complementary to the SMART III oligonucleotide, which then serves as a universal PCR priming site in subsequent amplifications.

MATERIALS AND METHODS

Preparation of Total RNA

Total RNA was isolated from LNCaP and HeLa cancer cell lines, derived from prostate and cervical cancers, respectively, and from normal and tumor colon tissues using the guani-

dinium isothiocynate (GITC) method and centrifugation through cesium chloride cushions (8).

SMART PCR cDNA Synthesis

Six first-strand cDNA reactions were performed from either spiked RNA samples or pure RNA derived from HeLa and LNCaP cell lines. HeLa RNA samples were spiked with LNCaP RNA amounting to 10% and 30% of the total. Conversely, the LNCaP RNA samples were spiked with 10% and 30% HeLa RNA. One microgram of total RNA was added to each RT reaction. The rest of the reagents used in the RT reaction were from the SMART PCR cDNA Synthesis Kit (CLONTECH). The manufacturer's instructions were followed during cDNA synthesis. The only modification was the addition of 40 U of RNasin® Ribonuclease Inhibitor (Promega, Madison, WI, USA) to each RT reaction. The first-strand cDNA reaction was diluted to 50 µL using 1× TE buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA). One microliter of the diluted cDNA from each RT reaction was used to generate PCR-amplified cDNA libraries using the primers and instructions in the kit, with minor modifications. Optimal PCR amplification conditions were determined following the instructions in the SMART PCR Synthesis Kit User's Manual (PT3041-1; CLONTECH) with slight modification. Ten-microliter aliquots of amplified cDNA samples removed after 15, 18, 21 and 24 cycles were resolved on 1× TAE (40 mM Tris-acetate, 1 mM EDTA, pH 8.3)/1.2% agarose gels and then transferred onto MSI-MagnaGraph® Nylon Membranes (Micron Separations, Inc. [MSI], Westborough, MA, USA). The membranes were probed with glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and transferrin receptor (TFR) cDNA to control for over-amplification. The optimum number of PCR cycles was determined to be one less than the number that gave the maximum yield of full-length GAPDH and TFR products in single discreet bands. The PCR amplification conditions were as follows: denaturation at 95°C for 1 min followed by 20 cycles of amplification (95°C for 10 s and 68°C for 5 min) in a Perkin-Elmer GeneAmp® PCR System 9600 (PE Biosystems, Foster City, CA, USA). All PCRs were set up in volumes of 100 µL. In a separate series of reactions, normal and tumor colon tissue SMART cDNA libraries were synthesized as described above. Approximately 1250–1750 µg of unpurified cDNA were typically obtained from amplifying the entire 50 µL RT reaction. Fifty to one hundred micrograms of purified cDNA were recovered following purification using the QIAquick™ PCR Purification Kit (Qiagen, Chatsworth, CA, USA).

Northern Hybridization

RNA was electrophoresed using a modification of the GITC/TBE gel method described by Goda and Minton (4). Ten-microgram aliquots of LNCaP and HeLa RNA were denatured in 2× sample buffer containing 20% formamide, 40% formaldehyde, 50 mM phosphate buffer (pH 7.0), 3 mM EDTA, 30% 5× Ficoll® blue solution (25% Ficoll 400, 0.1% EDTA, 0.5% sodium dodecyl sulfate [SDS] and 0.1% bromophenol blue) at 90°C for 5 min. Samples were fractionated by gel electrophoresis on a 5 mM GITC-1× TBE, 1.2% agarose gel. Gels were stained for 30 min in 1× TBE buffer (100 mM Tris, 90 mM boric acid, 1 mM EDTA) containing 0.5 µg/mL ethidium bromide and then destained for the same amount of time in 1× TBE. RNA was transferred onto MSI-MagnaGraph Nylon Membranes. RNA was cross-linked to the membrane in a UV-Stratalinker® 1800 (Stratagene, La Jolla, CA, USA). ^{32}P-labeled cDNA probes were generated using 50 ng of each cDNA insert in a random priming reaction consisting of 5 µL [α-^{32}P]dCTP (3000 Ci/mmol; NEN Life Science Products, Boston, MA, USA) and the reagents from an RT Random Priming Kit (Stratagene). Membranes were prehybridized for 1 h at 68°C in ExpressHyb™ solution (CLONTECH) and subsequently hybridized for 1 h at 68°C in ExpressHyb solution containing 10^6 cpm/mL of cDNA probe and 100 µg/mL salmon sperm DNA. Membranes were rinsed in 2× standard saline citrate (SSC), 0.05% SDS at room temperature for 30 min followed

by stringent washing in 0.1× SSC, 0.1% SDS at 60°C for 15 min. Radioactive probes were hybridized in succession to the same membrane, following stripping of the previous probe. Blots were exposed to Phosphor Imaging Screens (Molecular Dynamics, Sunnyvale, CA, USA) overnight.

cDNA Southern Hybridization

Two micrograms of unpurified cDNA from each of the six amplified HeLa, LNCaP and spiked libraries were fractionated on a 1.2% 1× TAE agarose gel. Six micrograms of purified normal and colon tumor cDNA were fractionated on a 5 mM GITC-1× TBE, 1.2% agarose gel along with 10 µg of the total RNA samples used to make the colon libraries. Gels were soaked in 0.25 *N* hydrochloric acid, neutralized in 0.5 M NaOH, 1.5 M NaCl and then soaked in 1 M Tris-HCl, pH 8.0, before being transferred onto MSI-MagnaGraph Nylon Membranes. DNA was cross-linked to the membrane in a UV-Stratalinker 1800. Probe preparation, membrane hybridization, membrane washing and analysis were carried out as described for northern hybridization.

Dot Blot Hybridization

cDNA inserts from S100P, maspin, prostate-specific antigen (PSA), TFR and GAPDH were obtained from recombinant plasmid DNA by restriction digestion and purification of inserts using the Wizard® PCR Preps DNA Purification Kit (Promega). The DNA was denatured for 5 min at room temperature in 0.1 M sodium hydroxide and neutralized in 0.4 M ammonium acetate. Twenty nanograms of each denatured cDNA insert were transferred in duplicate onto an MSI-MagnaGraph Nylon Membrane using a 96-well Minifold® Dot-Blot

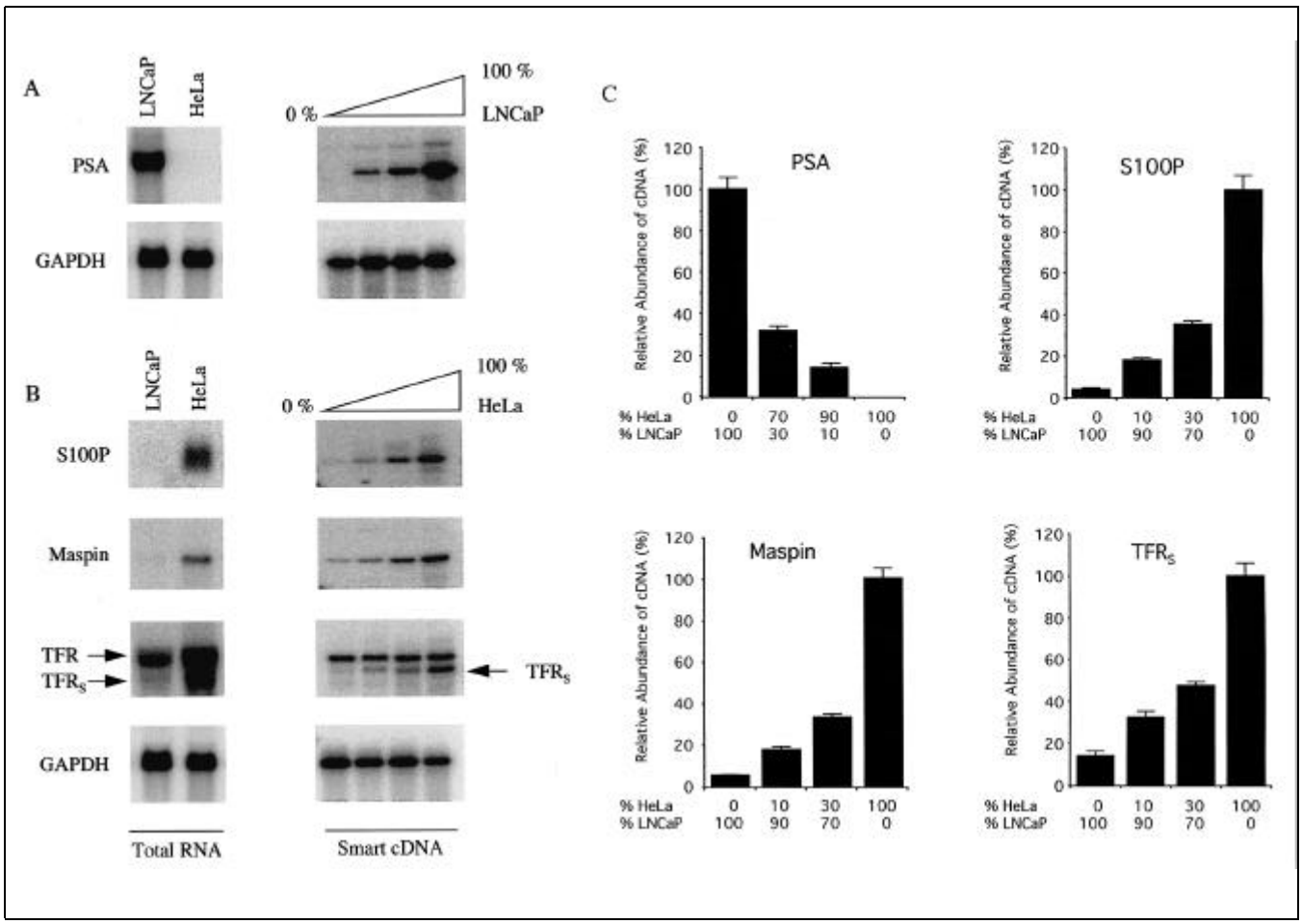

Figure 1. Maintenance of original message profiles in SMART PCR-amplified cDNA libraries. (A) Left panel: northern blot analysis of LNCaP and HeLa RNA showing the differential expression of the *PSA* gene in LNCaP cells. Right panel: cDNA Southern blot analysis of amplified cDNA libraries synthesized from HeLa RNA samples that were spiked with LNCaP RNA. The final percentage of LNCaP RNA was 0%, 10%, 30% or 100%. (B) Left panel: differential expression of *S100P*, *maspin* and TFR_S genes in RNA extracted from LNCaP and HeLa cells. Right panel: cDNA Southern blots obtained from amplified cDNA libraries synthesized from LNCaP RNA samples that were spiked with HeLa RNA. The final percentage of HeLa RNA was 0%, 10%, 30% or 100%. The northern blots contained 10 µg of total RNA per lane, while the cDNA Southern blots had 2 µg/lane of amplified cDNA. To control for variations in sample loading, the same blots were hybridized to a GAPDH probe. (C) Hybridization signals corresponding to the different amounts of PSA, S100P, maspin and TFR_S transcripts present in the amplified cDNA libraries were quantified from Phosphor Imaging screens using the ImageQuant™ software (Molecular Dynamics). The corresponding GAPDH signal was used to normalize for variations in cDNA loading. Error bars represent the standard deviation from 5 separate analyses of the 6 libraries.

Table 1. Summary of the Expression Profiles of 12 DD Amplification Products Analyzed by Northern and cDNA Southern Blotting

Probe	Differential Display	Northern	cDNA Southern
1	N	N	N
2	N	N	N
3	N	N	N
4	N	N	N
5	N	N	N
6	C	N	N
7	C	N	N
8	C	N	E
9	C	E	C
10	C	E	N
11	C	E	E
12	C	U	U

N = Over-expressed in normal colon sample.
C = Over-expressed in colon tumor sample.
E = Equally expressed in normal and tumor samples.
U = Undetectable expression level.

System (Schleicher & Schuell, Keene, NH, USA). The membranes were rinsed in 6× SSC and cross-linked in a UV-Stratalinker 1800. Membranes were prehybridized in ExpressHyb containing 100 µg/mL salmon sperm DNA. The cDNA used to generate random primed probes was synthesized using the SMART PCR cDNA Kit. Complex ^{32}P-labeled probes were made from 50 ng of SMART PCR-amplified LNCaP or HeLa cDNAs, using Stratagene's RT Random Priming Kit and 5 µL [α-^{32}P]dCTP (3000 Ci/mmol). Primers and dNTPs were removed from the amplified complex cDNA mixtures using the

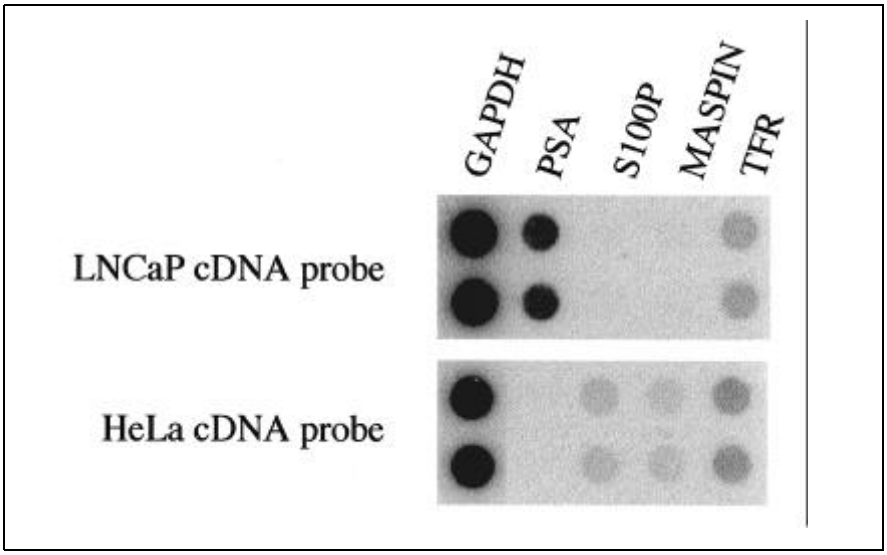

Figure 2. Detection of differential gene expression using SMART PCR-amplified cDNA as probes. LNCaP and HeLa SMART cDNA were synthesized as described in the text. 50 ng of each purified cDNA was labeled with ^{32}P[dCTP] using a RT Random Priming Kit to a specific activity of at least 10^6 cpm/ng and hybridized to dot blots containing 20 ng/spot of GAPDH, PSA, S100P, maspin and TFR.

Wizard PCR Preps Kit before using the cDNA for probe synthesis. Probes had specific activities of 10^6 cpm/ng or higher. Membranes were hybridized for 1 h at 68°C in 15 mL of ExpressHyb solution containing 3×10^6 cpm/mL of cDNA probe and 100 µg/mL salmon sperm DNA. Washes were performed as described for northern hybridization. Blots were exposed to Phosphor Imaging Screens for 48 h.

Differential Display

Differential display (DD) was performed using anchored primers AP1, AP2 and AP5 and arbitrary primers ARP13, ARP14, ARP15 and ARP16 supplied in the HIEROGLYPH™ mRNA Profile Kit 4 (Genomyx, Foster City, CA, USA). Manufacturer's instructions were followed during RT, DD-PCR, gel band recovery and re-amplification. Gel electrophoresis was performed using HR-1000™ Gel Reagents and a genomyxLR™ Programmable DNA Sequencer (both from Genomyx).

RESULTS AND DISCUSSION

To determine whether SMART cDNA libraries maintain the original message profile, six cDNA libraries were constructed either from pure HeLa or LNCaP RNA or from LNCaP RNA spiked with 10% and 30% HeLa RNA or HeLa RNA spiked with 10% and 30% LNCaP RNA. Purified cDNA fragments of five different genes were used as probes. Two of them, *maspin* and *S100P* were identified by DD (1,15). The others were *PSA*, *TFR* and *GAPDH*. Northern blot results (Figure 1, A and B, panels on the left) showed that PSA is a LNCaP-specific transcript, while the S100P and maspin transcripts are predominately expressed in HeLa cells. GAPDH and TFR transcripts are expressed at similar levels in both cell lines, while expression of TFR_S, a transcript that might be either a splice variant or a truncated form of the native TFR transcript, appears to be specific to HeLa cells. If the resulting cDNA libraries were representative, they should consist of 0%, 10%, 30% and 100% LNCaP-specific cDNAs in a background of HeLa cDNA (Figure 1A, right panels) and 0%, 10%, 30% and 100% HeLa-specific cDNAs in a background of LNCaP cDNA (Figure 1B, right panels). To verify this representation, the membranes were hybridized with the same probes as those used for the northern blots. Figure 1, A and B (panels on the right), shows the results from this cDNA Southern blot experiment. This experiment was repeated at least five times to determine if low- (S100P and maspin), medium- (PSA and TFR) and high- (GAPDH) abundance transcripts were amplified in a representative manner from the spiked cDNA samples. Figure 1C shows results from the five experiments, following normalization to GAPDH transcripts. These results show a good representative amplification of PSA from the LNCaP RNA that was used to spike HeLa RNA and the HeLa-predominant S100P, maspin and TFR_s transcripts when LNCaP RNA was spiked with HeLa RNA. GAPDH and the larger TFR transcript are present at equivalent levels in all libraries, accurately reflecting the equal expres-

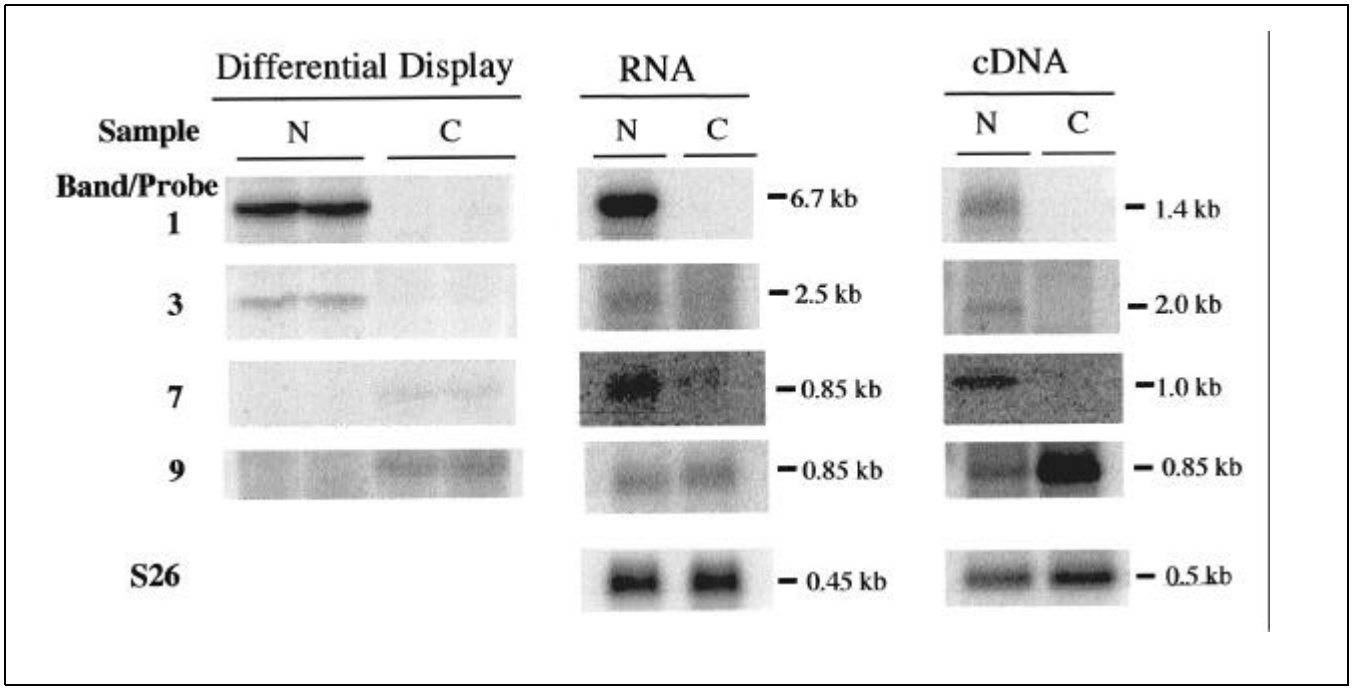

Figure 3. Verification of differential gene expression by northern blot analysis and cDNA Southern blot analysis using SMART PCR-amplified cDNA. Typical DD expression profiles, northern and cDNA Southern blot expression profiles obtained from normal (N) and tumor (C) colon RNA probed with radiolabeled differentially expressed amplification products identified using DD. The DD products were run in duplicate. The northern blots contained 10 µg of total RNA per lane, while the cDNA Southern blots had 6 µg/lane of amplified cDNA. Hybridization to a cDNA probe for the ribosomal protein S26 (12) was used to quantify variations in the amount of nucleic acid loaded. Included are representative blots showing that the levels of S26 in the normal and tumor samples varied by at most 2-fold.

sion in both cell lines. In the cases of maspin and S100P, the amplified cDNA library derived from 100% LNCaP RNA produces a detectable signal, while the northern blot does not. This is most likely due to increased sensitivity of the cDNA Southern, where 2 µg of unpurified cDNA (mRNA equivalent) were loaded per lane as compared with 10 µg of total RNA in the northern blot. Although 1 µg of total RNA was used to synthesize the cDNA described above, similar results were obtained when 50 ng of total RNA were used as the starting material for RT (data not shown).

In addition to demonstrating that SMART PCR cDNA libraries can be used to confirm differential gene expression on cDNA Southern blots, we wanted to determine whether they can be used as complex probes to confirm differential expression on dot blots. Others have previously reported using as much as 2 µg of mRNA or 5–40 µg of total RNA to synthesize amplified RNA probes for use in studying differential gene expression (10,11). In this report, we show that as little as 50 ng of our amplified cDNA can be used to perform similar experiments without compromising sensitivity for low-abundance transcripts. LNCaP-specific expression of PSA and HeLa-specific expression of S100P and maspin were confirmed. All cDNAs were easily detectable using the complex mixture of amplified cDNA probes, regardless of whether they represented very low- or high-abundance transcripts (Figure 2). While the spiking experiments described above demonstrated maintenance of the relative representation of a given transcript in multiple samples (horizontal representation), the dot blot results indicated that vertical representation was also maintained to some extent (maspin and S100P signals are lower than GAPDH, PSA and TFR signals). In a more quantitative study, 4600 arrayed genes were probed with labeled first-strand cDNA and random primed SMART PCR cDNA synthesized from the same total RNA sample. More than 75% of the arrayed genes generated signals that varied by less than twofold between the SMART PCR-amplified and the unamplified total RNA probes (Allison Berger, Millennium Predictive Medicine, Cambridge, MA, USA, personal communication).

Since the previous experiments utilized a model system of spiked RNA samples, we wanted to determine whether SMART cDNA libraries could be used to confirm differential expression of unknown genes identified by a commonly used gene discovery technology, DD (6). Typically, northern blots are used to confirm DD results, but the limiting amount of RNA in many clinical samples would favor the

use of amplified cDNA if representation can be maintained. DD was performed according to the manufacturer's instructions using a kit purchased from Genomyx. Twelve [α-^{32}P]dCTP-labeled amplification products obtained from DD gel bands were used to probe northern blots prepared with 10 µg of the normal colon RNA and colon tumor RNA used for DD. The same probes were hybridized to cDNA Southern blots prepared with 6 µg of SMART cDNA that was PCR-amplified from cDNA derived from the same normal colon and colon tumor RNA used for DD. Representative DD patterns are shown in the left panels of Figure 3. The center and right panels show the differential gene expression profile results obtained using standard northern blot analysis and SMART cDNA-amplified libraries, respectively. Table 1 summarizes the results from 12 genes identified as being differentially expressed by DD. Northern analysis confirmed five of these 12 differences. Although consistent with results of earlier DD experiments, the fraction of genes verified as being differentially expressed would most probably have increased if recent improvements to the technique had been used (7,9,14). Analysis using SMART cDNA-amplified libraries confirmed differential expression of the same five genes. Of the seven genes not confirmed by northern blots, six were also not confirmed by the SMART cDNA libraries. In one case (probe No. 9), SMART cDNA confirmed DD results when northern blots did not. Although PCR amplification is utilized in both the identification and verification processes, these results indicate that SMART cDNA might be an attractive substitute for total RNA to confirm DD patterns when RNA sources are limiting.

Here, we have shown that SMART PCR-amplified cDNAs can retain the representation of the original RNA. This was demonstrated in spiked RNA samples and in the verification of gene expression differences identified by DD. We have also shown that small amounts of the amplified cDNA libraries can be used as complex probes to confirm differential gene expression patterns of both rare and abundant transcripts on dot blots. The conditions we have described for SMART PCR cDNA amplification have been optimized for the Perkin-Elmer GeneAmp System 9600. We have found that these amplification parameters give consistent results when samples were derived from either cell lines or human tissue. From the agarose gel analyses results, it is reasonable to conclude that SMART PCR cDNA technology is able to amplify full-length transcripts of *TFR* (ca. 5.5 kb), *maspin* (ca. 2.8 kb), *PSA* (ca. 1.5 kb), *GAPDH* (ca. 1.4 kb), *S100P* (ca. 600 bp), *S26* (ca. 500 bp) and fifteen other genes whose transcripts range in size from 0.9–2.5 kb (Figure 1, A and B, Figure 3A and results not shown). In a few cases, however, it appears that the SMART PCR cDNA procedure does not generate full-length cDNA, as evidenced by the difference in migration of the RNA and cDNA that hybridized to probe No. 1 in Figure 3.

ACKNOWLEDGMENTS

W.O. Endege and K.E. Steinmann contributed equally to this paper. We thank the Pardee laboratory at The Dana Farber Cancer Institute, Boston, MA for supplying the PSA, maspin and S100P plasmids. We thank Dr. Doug Ludtke and Wallace Nduati at Chiron Diagnostics for providing their 5 mM GITC-1× TBE, 1.2% agarose gel protocol. We thank Allison Berger at Millennium Predictive Medicine, Cambridge MA, for sharing her unpublished data on the vertical representation of SMART PCR cDNA.

REFERENCES

1. **Averboukh, L., P. Liang, P.W. Kantoff and A.B. Pardee.** 1996. Regulation of S100P expression by androgen. Prostate *29*:350-355.
2. **Brady, G. and N.N. Iscove.** 1993. Construction of cDNA libraries from single cells. Methods Enzymol. *225*:611-623.
3. **Froussard, P.** 1992. A random-PCR method (rPCR) to construct whole cDNA library from low amounts of RNA. Nucleic Acids Res. *20*:2900.
4. **Goda, S.K. and N.P. Minton.** 1995. A simple procedure for gel electrophoresis and Northern blotting of RNA. Nucleic Acids Res. *23*:3357-3358.
5. **Klickstein, L.B., R.L. Neve, E.A. Golemis and J. Gyuris.** 1996. Conversion of mRNA into double-stranded cDNA, p. 5.5.2-5.5.14. *In* F.M. Ausubel, R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith and K. Struhl (Eds.), Current Protocols in Molecular Biology, Vol. 1. John Wiley & Sons, New York.
6. **Liang, P. and A.B. Pardee.** 1992. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. Science *257*:967-971.
7. **Miele, G., L. MacRae, D. McBride, J. Manson and M. Clinton.** 1998. Elimination of false positives generated through PCR re-amplification of differential display cDNA. BioTechniques *25*:138-144.
8. **Maniatis, T., E.F. Fritsch and J. Sambrook.** 1982. Molecular Cloning: A Laboratory Manual. CSH Laboratory Press, Cold Spring Harbor, NY.
9. **Martin, K.J., C.-P. Kwan, M.J. O'Hare, A.B. Pardee and R. Sager.** 1998. Identification and verification of differential display cDNAs using gene-specific primers and hybridization arrays. BioTechniques *24*:1018-1026.
10. **Poirier, G.M.-C., J. Pyati, J.S. Wan and M.G. Erlander.** 1997. Screening differentially expressed cDNA clones obtained by differential display using amplified RNA. Nucleic Acids Res. *25*:913-914.
11. **Van Gelder, R.N., M.E. von Zastrow, A. Yool, W.C. Dement, J.D. Barchas and J.H. Eberwine**. 1990. Amplified RNA synthesized from limited quantities of heterogeneous cDNA. Proc. Natl. Acad. Sci. USA *87*:1663-1667.
12. **Vincent, S., L. Marty and P. Fort.** 1993. S26 ribosomal protein RNA: an invariant control for gene regulation experiments in eukaryotic cells and tissues. Nucleic Acids Res. *21*:1498.
13. **Zhang, H., R. Zhang and P. Liang.** 1996. Differential screening of gene expression difference enriched by differential display. Nucleic Acids Res. *24*:2454-2455.
14. **Zhao, S., S.L. Ooi, F.-C. Yang and A.B. Pardee.** 1996. Three methods for identification of true positive cloned cDNA fragments in differential display. BioTechniques *20*:400-404.
15. **Zou, Z., A. Anisowicz, M.J. Hendrix, A. Thor, M. Neveu, S. Sheng, K. Rafidi, E. Seftor and R. Sager.** 1994. Maspin, a serpin with tumor-suppressing activity in human mammary epithelial cells. Science *263*:526-529.

Received 16 July 1998; accepted 16 December 1998.

Address correspondence to:

Dr. Wilson O. Endege
Millennium Predictive Medicine, Inc.
Building 700, 1 Kendall Square,
Cambridge, MA 02139-1562, USA
Internet: endege@mpi.com