Available online at www.sciencedirect.com
SCIENCE DIRECT
Biomolecular Engineering 23 (2006) 41–54
Biomolecular Engineering
www.elsevier.com/locate/geneanabioeng

Targeted cDNA differential display (TcDD)

Giang H. Nguyen [a], Nam C. Nguyen [a], Niels Storm [a], Chris Gange [a], Haralambos Gavras [b], Cassandra L. Smith [a,*]

[a] *Molecular Biology Research Laboratory and Departments of Biomedical Engineering, Biology, and Pharmacology and Experimental Therapeutics, 36 Cummington Street, Boston, MA 02215, United States*
[b] *Department of Medicine, Boston University School of Medicine, Boston University, 800 East Concord, Boston, MA, United States*

Received 5 September 2005; received in revised form 3 December 2005; accepted 5 December 2005

Abstract

Targeted cDNA differential display (TcDD) was developed to study expression of a different selected gene families especially those at low copy numbers per cell. This method is an adaptation of our previously described targeted genomic differential display method (TGDD). In TcDD, the expression of genes containing target sequences such as CAG repeating sequences or genes encoding for zinc-finger binding proteins were followed in an experimental rat model with salt-induced hypertension. DNA sequencing experiments demonstrated that the effectiveness of targeting was greater than 99%.

Keywords: Differential display; cDNA; Gene expression; Gene families; Repeats

1. Introduction

In mammalian cells, total RNA is made up of 96% rRNA and ~4% mRNA (Alberts, 1994; Bonaldo et al., 1996; Zhang et al., 1997). Ninety-five percent of mRNAs are present at ≤5 copies per cell and represent 35–50% of the total mRNA mass. Rare mRNAs (<1 per cell) mostly occur in the brain, and account for the majority of mRNAs (Chaudhari and Hahn, 1983; Hastie and Bishop, 1976; Kaplan et al., 1978; Young and Birnie, 1976). Many of these rare transcripts are involved in complex signaling pathways. This means that expression studies on brain, as well as other tissue, need to resolve 50,000 different transcripts and detect 1 in 500,000 mRNAs in order to detect all genes.

A large number of protocols have been developed to monitor global gene expression at the transcription level. However, no one method will provide all differences between samples (For reviews see Smith et al., 2000; Liang and Pardee, 1997; Bowtell and Sambrook, 2002; Lorkowski and Cullen, 2003; Sturtevant, 2000; Ali et al., 2001). Briefly, there are two basic approaches to monitoring mRNA expression: subtractive hybridization (SH) and display methods. SH experiments output large (≥5000 base pairs (bp)) differences between paired samples. Although a large number of subtractive methods have been described, the principal method currently in use is based on PCR suppression (see below) described in Siebert et al. (1995).

In the display approaches, a complex fingerprint is created from pools of mRNAs, or more commonly, from their DNA copies (cDNAs). This fingerprint is created by electrophoretic size fractionations, or by sequence fractionation in hybridization experiments (aka DNA arrays). Potentially, DNA arrays can analyze up to 40,000 different transcripts (e.g. http://www.affymetrix.com). In practice, however, each array experiment produces usable data for only about half of these transcripts, and is limited to detecting only 1 in 300,000 transcripts.

The differential display (DD) protocol that was developed by Liang and Pardee (1992) randomly amplified and radiolabeled groups of cDNAs that were size fractionated by high-resolution polyacrylamide electrophoresis. The cDNAs were made by reverse transcriptase of mRNA using a oligo T primer. This is followed by PCR amplification using the same oligo T primer in combination with a random primer. Currently, several thousand publications have used this protocol. In DD, mRNA species are identified by size, and differences in intensities of specific bands highlight differences in gene

* Corresponding author. Tel.: +1 617 353 8503; fax: +1 617 353 8506.
E-mail address: clsmith@bu.edu (C.L. Smith).

1389-0344/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.bioeng.2005.12.001

expression between compared samples. Specific differences must be cloned, sequenced, and individually confirmed.

An alternative DD approach is based on cloning and sequencing bits of each mRNA from a cell. This approach was performed on a large scale first by Incyte (http://www.incyte.com). Another approach is called serial analysis of gene expression (SAGE; http://www.sagenet.org, Velculescu et al., 1995; Vilain and Vassart, 2004). SAGE is also based on sequencing but has greater through put than the Incyte approach. In SAGE, DNA sequencing is performed on DNA molecules composed of ligated (~20-mer) sequences from random mRNAs. As originally described, each DNA sequencing reaction yields information on 200 transcripts rather than one (or 20,000 transcripts per 100 sequencing runs). The downside is that SAGE requires larger amounts of mRNA than DD methods, and is more labor intensive. Recent SAGE enhancements (e.g. miniSAGE, microSAGE or SADE (SAGE adaptation for downsized extracts)) allow small amounts of material to be analyzed (summarized in Lorkowski and Cullen, 2003). Gene expression differences detected with any method need to be confirmed by testing using Northern blots or quantitative PCR.

All of the methods described above preferentially provide information about genes expressed at high levels. In the future, DNA sequencing improvements should allow more transcripts to be analyzed per sample, creating "deeper" expression libraries that allow for the analysis of genes expressed at low levels.

Targeted gene analysis is an alternative to random gene expression analysis. Targeted gene analysis examines the expression level of a selected set of genes. This approach allows studies to focus on genes expressed at low levels.

In previous studies, we described protocols for targeted genomic differential display (TGDD) (Broude et al., 1997, 1999; Bouchard et al., 1999). TGDD was used to compare the genomes of twins afflicted with schizophrenia (Nguyen et al., 2003; Lavrentieva et al., 1999) and to develop simple repeat markers for Trypanosomes (Oliveira et al., 1998). Random sequencing of the selected DNA pools in the TGDD experiments revealed greater than 99% enrichment for genomic fragments containing the target sequence. Here, the principles developed in the TGDD experiments were applied to mRNA analysis. The adapted method is called targeted cDNA differential display (TcDD).

2. Material and methods

2.1. Samples

The synthetic oligonucleotides were obtained from Operon (http://www.operon.com) unless otherwise indicated (Table 1). Tissue samples were obtained from adult male Wistar rats used in an animal model of hypertension as described by Hatzinikolaou et al. (1980). In this model system, the right kidney was removed 2 weeks prior to the removal of the left kidney. Two hours after the last surgery, the anephric rats were infused with 2 ml of 4%, 8% NaCl or 5% dextrose for 2 hours before the animals were sacrificed. Heart, aorta, medulla/pons, cerebellum were harvested immediately and homogenized in the "Lysis Buffer" of the Qiagen (http://www.qiagen.com) "Direct mRNA Isolation Kit". The samples were stored at −70 °C. RNA isolation was performed following Qiagen's directions. The final eluate containing RNA was ethanol precipitated and re-dissolved in 5 μl water. cDNA was synthesized using reagents from the Clontech PCR Select cDNA Subtraction Kit that included an oligo dT primer. In later experiments, AMV reverse transcriptase was obtained from New England Biolabs or Roche.

Table 1
Summary of animal samples

Animal	Treatment	MAP (mm Hg)[a]	
		Before	After
1	4% NaCl	118	136
2	8% NaCl	120	160
3	5% dextrose	102	100

For each animal, mRNA was purified from heart, thoracic aorta, medullar/pons and cerebellum immediately after treatment with the indicated infusion. Animals were infused with hypertonic saline or iso-osmotic dextrose over a 2 h period, receiving a volume of 2 ml. They were sacrificed on hour after the discontinuation of the infusion. The mean arterial pressure (MAP) did not change during this interval.

[a] MAP ~2 h after right kidney removal and before start of the infusion (before) or after the 2 h infusion (after).

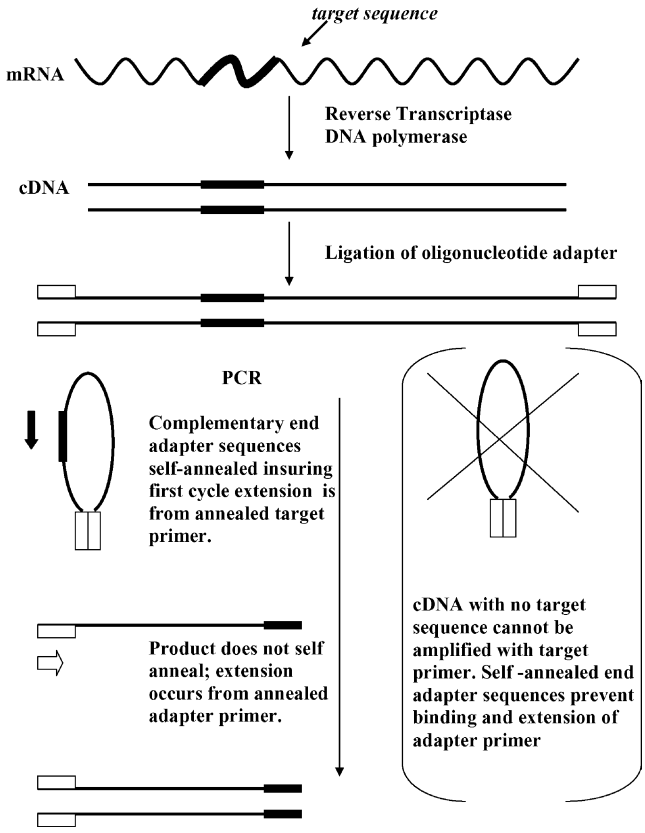

Fig. 1. Schematic of TcDD protocol. TcDD first converted mRNA to cDNA using conventional methods. The cDNAs were end tagged by ligation with a long 40 bp adapter sequence that could sever as PCR primer targets. The first PCR cycle initiated from the target primer within a single-stranded region of the cDNA because the complementary end-adapters annealed to each other and prevented binding and extension of the adapter primer. The second PCR cycle initiated with the annealing of the adapter to the single-stranded complementary end sequence.

2.2. TcDD (summarized in Fig. 1)

The double stranded cDNA was cleaved with the restriction enzymes *Hae*III (usually 20 U per 500 ng DNA). The sample was phenol extracted, ethanol precipitated, and re-dissolved in 9 μl of water. The adapters (*Hae*III adapter 43 and *Hae*II adapter 11 (numbers 1 and 2 in Table 2, respectively)) were ligated to the restriction-cleaved DNA as described in Bouchard et al. (1999) in a 30 μl reaction composed of 19 μl of cDNA, 2 U (in 2 μl) T4 DNA ligase (Epicentre Technologies, Madison, WI) and 20 μM each of oligonucleotides 1 and 2 (Table 2). The reaction was incubated at 16 °C overnight. Excess adapters were removed using the Qiagen PCR Purification Kit and recovery was maximized by sequential elution with 50 and 30 μl buffer, respectively. One of the most critical steps for TcDD is determining the DNA concentrations. The final DNA concentration was most reliably determined on ethidium bromide stained gels by comparison to a standard.

The TcDD protocol consisted of two semi-nested consecutive hot start PCRs. Usually, the first 25 μl PCR contained 10–12 ng prepared DNA, 5 mM $MgCl_3$, 0.25 mM of each dNTP, 5 pM of each primer, and 1.25 U of enzyme (see below). The products of the first PCR were diluted 1/100 and used as templates for a second PCR. Usually conditions for the second PCR were the same as in the first PCR. The first PCR used the adapter primer NA21 (oligonucleotide 3 in Table 2) along with a targeting primer. Experiments targeting CAG-repeated sequences used oligonucleotides 8–11 in Table 2. In some experiments, oligonucleotide 8 was used in the first PCR and oligonucleotide 11 in the second PCR. Oligonucleotide 12 (Table 2) was used in experiments targeting zinc-finger gene sequences.

Hot start PCR was insured by the addition of antiTaq antibody from Clontech (http://www.bdbiosciences.com/clontech). Here, a mixture of AmpliTaq DNA polymerase and KlenTaq1 (Perkin-Elmer) or HIFI enzyme mix (Roche) were used. The reactions were mixed, centrifuged and subjected to an initial denaturing at 94 °C for 2 min, 25 cycles of 94 °C for 15 s, 70 °C for 30 s, 72 °C for 60 s and final extension step for 7 min at 72 °C in Peltier-Effect Cycling (PTC100, MJ Research; http://www.mjr.com). Usually, 20 rather than 25 cycles were used in the second PCR.

Most times the PCR products were analyzed on a 2% agarose gel before fractionation on a high-resolution DNA sequencing gel. Usually a 2.5 μl sample from the second PCR was denatured for 5 min at 94 °C in 3.5 μl of stop solution containing 6 mg/ml dextran blue and 0.1% sodium dodecyl sulfate in deionized formamide (Amersham, Pharmacia Biotech). Then, the sample was quickly cooled and loaded on to a 6% denaturing polyacrylamide gel in 0.6X TBE and electrophoresed on the ALF-Express Sequencer (Pharmacia Biotech) as recommended by the manufacturer. A Cy-5 labeled 50 base pair (bp) ladder (50–500 bp; Pharmacia Biotech) was used as a size standard. The results were visualized and compared using the Fragment Manager Software provided with the instrument.

2.3. DNA sequencing

The products of the second PCR was fractionated on 3% Metaphor or 5% (3:1 (w:w)) NuSieve:LE Agarose (FMC BioProducts, Rockland, ME) using 1X TAE buffer. Specific bands were identified by visualization after ethidium bromide staining on a UV transilluminator. A P10 Pipetman tip was used to ''pick'' pieces of agarose containing specific fragments. The agar was added to a PCR using the conditions described for the second PCR above. This procedure was repeated 2–3 times. An aliquot of the final re-amplification was diluted 100-fold and amplified in two 100 μl PCRs to produce sequencing template using an unlabeled target primer. The two reactions were combined, phenol extracted, ethanol precipitated and re-dissolved in 20 μl water. In some cases, the amplification products were separated on a 3% Nusieve: LE (3:1 (w:w)) agarose gel (FMC BioProducts). The band of interest was isolated using Qiagen's Gel Extraction Kit, eluted, lyophilized and dissolved in 10–15 μl of water. Cycle sequencing was performed using the Amersham-Pharmacia protocol with dNTP, ddNTP, dimethyl sulfoxide, and a Cy5 labeled primer in a final reaction volume of 20 μl overlaid with 20 μl of oil and incubations at 94 °C for 2 min, 36 cycles of 94 °C for 15 s, 65 °C for 15 s, and 72 °C for 40 s, and a final incubation at 72 °C for 10 min. Then, 4 μl of stop solution was added to the reaction before loading the entire sample on to ALF-Express sequencing gel. Usually, 5 μl contains 0.25–0.5 μg. Better sequencing results were obtained using ST19 or

Table 2
Oligonucleotides used in this work

Number	Name	Sequence (5′ → 3′)[a,b]	T_m (C)[c]
Adapters			
1	*Hae*III adapter 43	TGTAGCGTGAAGACGACAGAAA	
		GGGCGTGGTGCGGAGGGCGGT	
2	*Hae*III adapter 11	ACCGCCCTCCG	
Adapter-primers			
3	NA21	TGTAGCGTGAAGACGACAGAA	55
4	ST19	CGTGGTGCGGAGGGCGGT	59
5	ST19Seq	GCGTGGTGCGGAGGGCGGTAC	61
6	ST19-Hae-AC	GTGCGGAGGGCGGTCCac	59
7	ST19-Hae-TG	GTGCGGAGGGCGGTCCtg	58
Target-primers			
8	$(CTG)_6$A	$(CTG)_6$a	59
9	$(CTG)_6$T	$(CTG)_6$t	59
10	$(CTG)_6$G	$(CTG)_6$g	60
11	$(CTG)_6$AC	$(CTG)_6$ac	60
12	Znf-forw	CWGMGGAYKCAYACDGGRGARAARCC[d]	50–63[e]

[a] Lower case letters indicate the unique anchor nucleotides adjacent to the conserved sequences of the adapter sequence, residual restriction enzyme recognition site, and target sequence.

[b] The standard IACUC nomenclature is used for the degenerate nucleotide positions in the zinc-finger primers.

[c] T_ms were calculated using the nearest neighbor method described by Breslauer et al. (1986) using http://www.basic.nwu.edu/biotoools/oligocalc.html).

[d] The zinc targeting primers were adapted from (Becker et al., 1995). The number of unique sequence present in each synthesis is a function of the number of the degenerate nucleotides. In these experiments, 0.8 mM concentrations of the Znf- had an effective unique primer concentration of 1.4 nM when corrected for the 768 unique sequences present. Standard IUPAC nomenclature is used for ambiguous bases: (M (A, C), R (A, G), W (A, T), S (C, G), Y (C, T), K (G, T), V (A, C, G), H (A, C, T), D (A, G, T), B (C, G, T), N (A, G, C, T).

[e] The range of T_m reflects the possible sequence variations of individual zinc-finger primers.

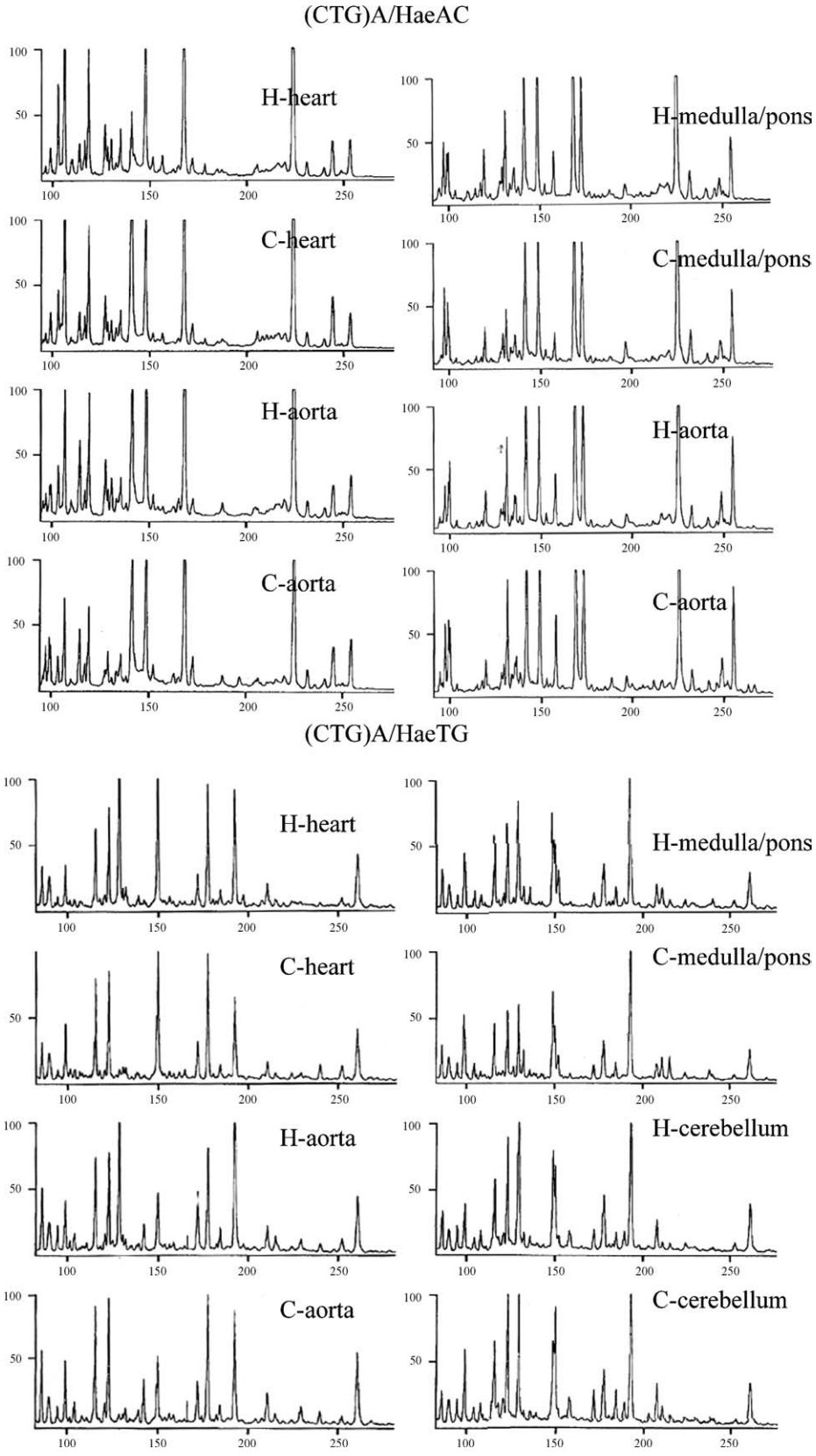

Fig. 2. Comparison of $(CAG)_n$ encoding mRNAs. cDNAs were obtained from a hypertensive (H) and a control (C) animal. TcDD analyses were done with a $(CAG)_6$ target primer with a 3′ end anchor sequence (A in the first PCR and AC in the second PCR (oligonucleotides 8 and 11, Table 2) and a adapter primer with a 3′ end AC or TG anchor sequence (oligonucleotides 6 or 7, Table 2). The organ from which the mRNA sample was obtained is indicated.

ST19Seq (oligonucleotides 4 and 5, respectively, in Table 2) rather than Znf-Forw (oligonucleotide 12 in Table 2) as sequencing primers.

3. Results

The major goal of this work was to develop a general method for analyzing selected mRNA gene families expressed at low levels. The mRNA samples were isolated by conventional methods from Wistar rats with salt-dependent hypertension that were induced by nephrectomy and saline load to mimic end-stage renal disease as described by Hatzinikolaou et al. (1980). cDNA templates for TcDD were made in a conventional reverse transcriptase reaction using an oligoT primer.

The principles of TcDD (Fig. 1) are similar to those developed for TGDD (for review, see Bouchard et al., 1999). TGDD uses PCR, sometimes preceded by hybridization, to create pools of genomic DNA fragments containing a selected target sequence. Here, TGDD was adapted to TcDD using PCR alone. These experiments were designed to work with small (~10 ng) amounts of sample because of the limited quantities of samples that are available for analysis in most experiments. The single most critical factor for reproducible TcDD was the determination of nucleic acid concentrations.

In TGDD and TcDD, sample complexity is reduced and the study focuses on a group of fragments containing a selected sequence. Selection is achieved through PCR using a target primer and PCR suppression. In PCR suppression, long (~40-mer) GC-rich oligonucleotides (called adapters) are ligated to the ends of restriction fragments. A target primer and a primer to the adapter sequences are used in the PCR. The adapter and/or target primers may have ''anchor'' nucleotides that anneal to unique sequences adjacent to the constant sequences (e.g. oligonucleotides 6 and 7, and oligonucleotides 8–11, respectively, Table 2). The anchor sequences reduce the number of fragments that are amplified. In these approaches, the first cycle of the PCR must initiate from the target sequences because the long adapters prevent the hybridization and extension of the adapter primer.

3.1. TcDD targeting of $(CAG)_n$ sequences

The adaptation of the TGDD protocol to TcDD involved the generation of a cDNA library by conventional methods, then cleavage with a restriction enzyme, ligation of adapters and the same PCR as described for TGDD. The cDNA was prepared for TcDD in the same manner as genomic DNA is prepared for TGDD. The cDNA was cleaved with a restriction enzyme, ligated to adapters, and used as template in a targeting PCR reaction using suppression hybridization as described above. To date, most TGDD experiments targeted sequences containing $(CAG)_n$ because of the importance of expanded CAG repeats in an increasing number of neurological diseases (for review see Cummings and Zoghbi, 2000), and because these sequences preferentially occur in coding sequences (Stallings, 1994). An example of TcDD experiments targeting $(CAG)_n$ encoding mRNA is shown in Fig. 2. The method's reproducibility is demonstrated by the similarity of the mRNA signals obtained from different tissues of the same animals. In these experiments, the $(CAG)_n$ primer is a 3′AC anchor (oligonucleotide 11, Table 2) and the adapter primer had a 3′ dinucleotide (5′-AC-3′ or 5′-TG-3′) anchor (oligonucleotides 6 and 7 Table 2). As expected, different $(CAG)_n$ containing fragments were amplified when the anchor sequence was varied.

3.2. TcDD targeting of zing-finger motifs

TcDD experiments targeting zinc-finger encoding cDNAs are shown in Figs. 3–6. The major step for developing a targeting protocol for a specific gene family is titrating the primers used in the PCR. This titration is performed to correct for differences in the number of target sites and the melting temperatures of the primers. For instance, all fragments ($\sim 10^6$) have the adapter primer site, while only a small subset of the fragments (~1600 – estimated from Ravasi et al., 2003) will also have a target site for the zinc-finger primer sequence (oligonucleotide 12, Table 2). Furthermore, the zinc-finger primer used in these experiments had many ambiguous base sites. The T_ms for individual primers ranged from 50 to 63 C (Table 1). This means that the sequences were not likely to be amplified evenly.

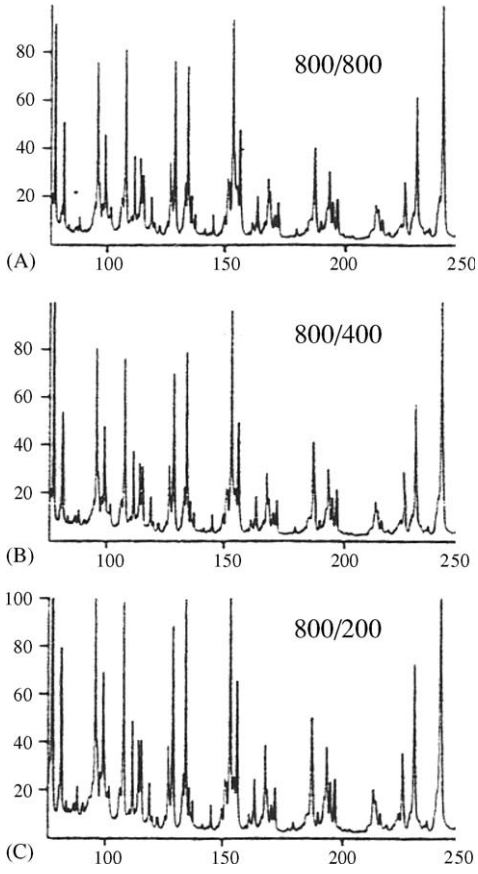

Fig. 3. Reproducibility of TcDD using different amounts of adapter primer. In these experiments, the Zn-forw primer (oligonucleotide 12, Table 2) was kept constant at 800 nM in both PCRs, the NA21 primer (oligonucleotide 3, Table 2) was kept constant at 800 nM in PCR I and the A-primer ST19 (oligonucleotide 4, Table 2) in the second PCR (see Section 2) varied between (A) 800 nM, (B) 400 nM and (C) 200 nM.

Fig. 3 shows a portion of a primer titration experiment. Nine reactions were performed varying the concentration of each primer independently (800, 400 or 200 nM). Note that the T_m will vary by a function of the concentration of short DNAs (Breslauer et al., 1986) therefore titration of each primer allows for adjustments in T_m.

The largest number of clearly defined fragments and the greatest yields were observed when the Zn-forw primer was present at 800 nM; minor differences were observed when the ST19 primer concentration was varied (Fig. 3). It should be noted that the effective concentration of a single Zn-forw primer sequence was 800 nM/786 = 1.0 nM when degeneracy of the primer is taken into account. A second optimization procedure was performed varying the number of PCR cycles (Fig. 4). Figs. 5 and 6 show TcDD results targeting Zn-finger encoding mRNAs from various tissues of animals with different infusions. The results demonstrate the reproducibility of the method. The results in Fig. 1 also demonstrate this reproducibility.

Earlier TGDD experiments demonstrated that >95% effectiveness in targeting of $(CAG)_n$ repeat sequences (Broude et al., 1999). In this study, similar experiments were performed with eight randomly chosen Zn-finger encoding fragments (Table 3). Seven of the clones had inserts while one did not. The obtained sequences minus the primer sequence were compared to rat and to all mammalian sequences in Genbank using megaBLAST. The primer sequence was eliminated in the BLAST analysis to prevent inappropriate biases toward sequences that contain the target. The results showed that all sequences had high identity to known rat or other mammalian zinc-finger encoding sequences. Furthermore, the target primer sequence was at the expected position on the genomic sequences in Genbank. None of the mismatches occurs at the 3′ end of the target primer. As the Wistar rat genome is not known, the results demonstrate targeting was ∼100% effective, despite the presence of a large number (9/26) of ambiguous bases in the zinc-finger target primer.

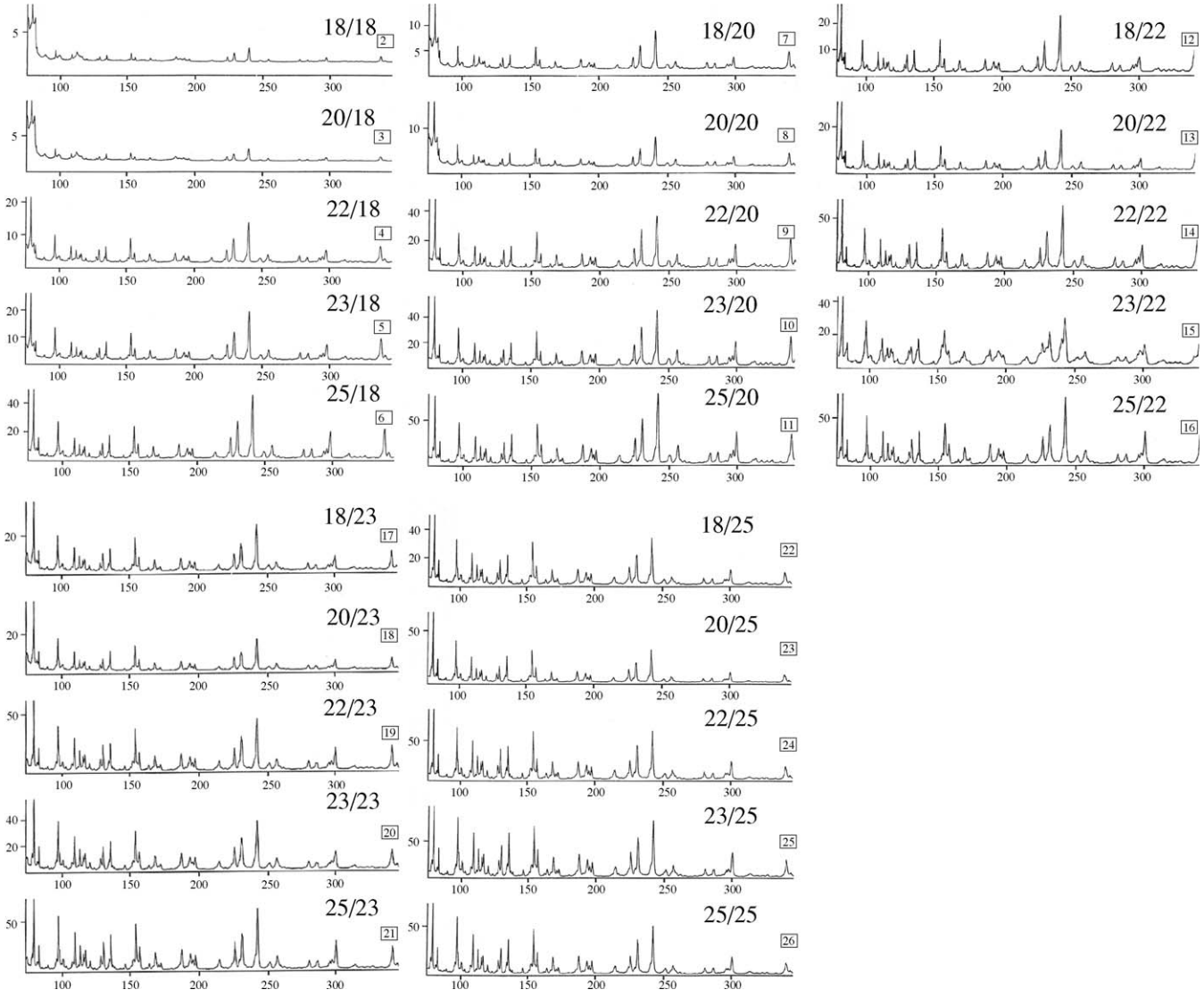

Fig. 4. Effect of cycle number on TcDD of zinc-finger encoding transcripts. In these experiments, 200 nM Na21, 200 nM ST19, and 800 nM Znf-forw primer oligonucleotides 3 and 12 (Table 2) in PCR I and oligonucleotides 4 and 12 (Table 2) with 3.5 mM $MgCl_2$ and DNA from animal number 3 (Table 1) were used.

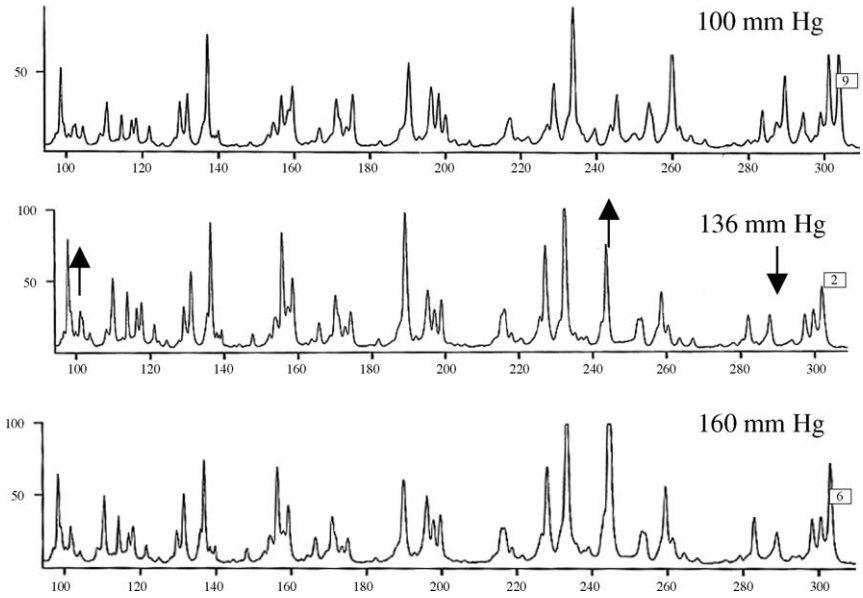

Fig. 5. Comparison of genes encoding zinc-finger mRNA expression levels in the heart of animals perfused with NaCl using TcDD. In these TcDD experiments, both PCRs used the targeting primer Znf-forw adapter combinations described in Fig. 4 and 25 cycles. The first PCR used 9.4 U KlenTaq and 3.5 mM $MgCl_2$. The second PCR used 1.25 U AmpliTaq and 5 mM of $MgCl_2$. Shown are the unprocessed fluorescent intensity measurements plotted as a function of time (min = ∼bp size).

4. Discussion

The focus of these experiments was to demonstrate the effectiveness of the TcDD procedure. Two target sequences were studied, a simple repeat sequence and a zinc-finger motif. Simple repeats targeting require the use of unique anchor sequences to insure that the target primer anneals to an end of the repeating sequences. Otherwise, the target primer can anneal anywhere along the repeat and several products will be amplified from each repeat occurrence. In this study, the experiments the target primer contained anchor sequences at the 3′ end. The anchor nucleotides (unique sequences adjacent to the target sequence or the restriction enzyme) reduce the target pool and select a subset of the repeat containing sequences. The selection of a subclass based on an anchor will become more sophisticated as more sequence and functional information becomes available.

A more complex nucleotide gene family motif is best created from a comparison of coding sequences. However, most protein motifs are developed from amino acid comparisons. The conserved amino acid motifs can then be translated back into a consensus coding sequence. This approach may introduce unnecessary variability and lower specificity in the target sequence however, because most amino acids are encoded by more than one codon. In this study, the zinc-finger target primer was based on a commonly used human consensus sequence described in Becker et al. (1995). The primers contained part of the zinc-finger consensus sequence. The primer included the sequence coding for two histidines plus the in between linkage regions, but did not include the coding regions for the two cysteines. The ambiguous bases were used because of the redundancy in the amino acid code as well as the fact that one consensus position contains multiple amino acids. Although this primer was effective at amplifying a large number of transcripts, a better understanding of the zinc-finger family proteins in rats should provide an improved framework for designing primers with fewer ambiguous bases. In general, the increasing generation of sequences and the accompanied analysis will provide an extensive framework to develop many targeting motifs useful for TcDD and TGDD. These motifs may even be species specific.

For example, one computational study (number 2 in Table 4), developed 30 primer octanucleotide sequence pairs that are present in ∼75% of human gene coding sequences. The same study identified 20 primer octanucleotide sequences that were present in ∼70% of human G-protein coupled receptor gene sequences. Cloning and sequencing of 30 PCR products using the G-protein coupled receptor sequence revealed a ∼24% success rate.

Targeting PCR based methods developed by other groups and used for gene expression studies are summarized in Table 4. Most approaches are based on DD or on RNA arbitrary primer (RAP). In the original DD methods, oligoT and experimentally optimized selected random sequences (10 and 20-mers) are used to reverse transcribe and amplify pools of cDNAs (Liang and Pardee, 1992). In RAP, two random primers, again empirically chosen, are used to amplify portions of cDNAs.

Targeting can be achieved in both the reverse transcriptase step and the PCR amplification step, or only in the PCR amplification step. The former approach has the advantage that the cDNA samples can be analyzed for many gene families. In this study, the cDNA samples are created using an oligoT and a random (or semi-random) primer for the reverse transcription reaction. When a reverse transcriptase reaction uses a targeting primer, the PCR amplification reaction uses the same primer, or an overlapping one.

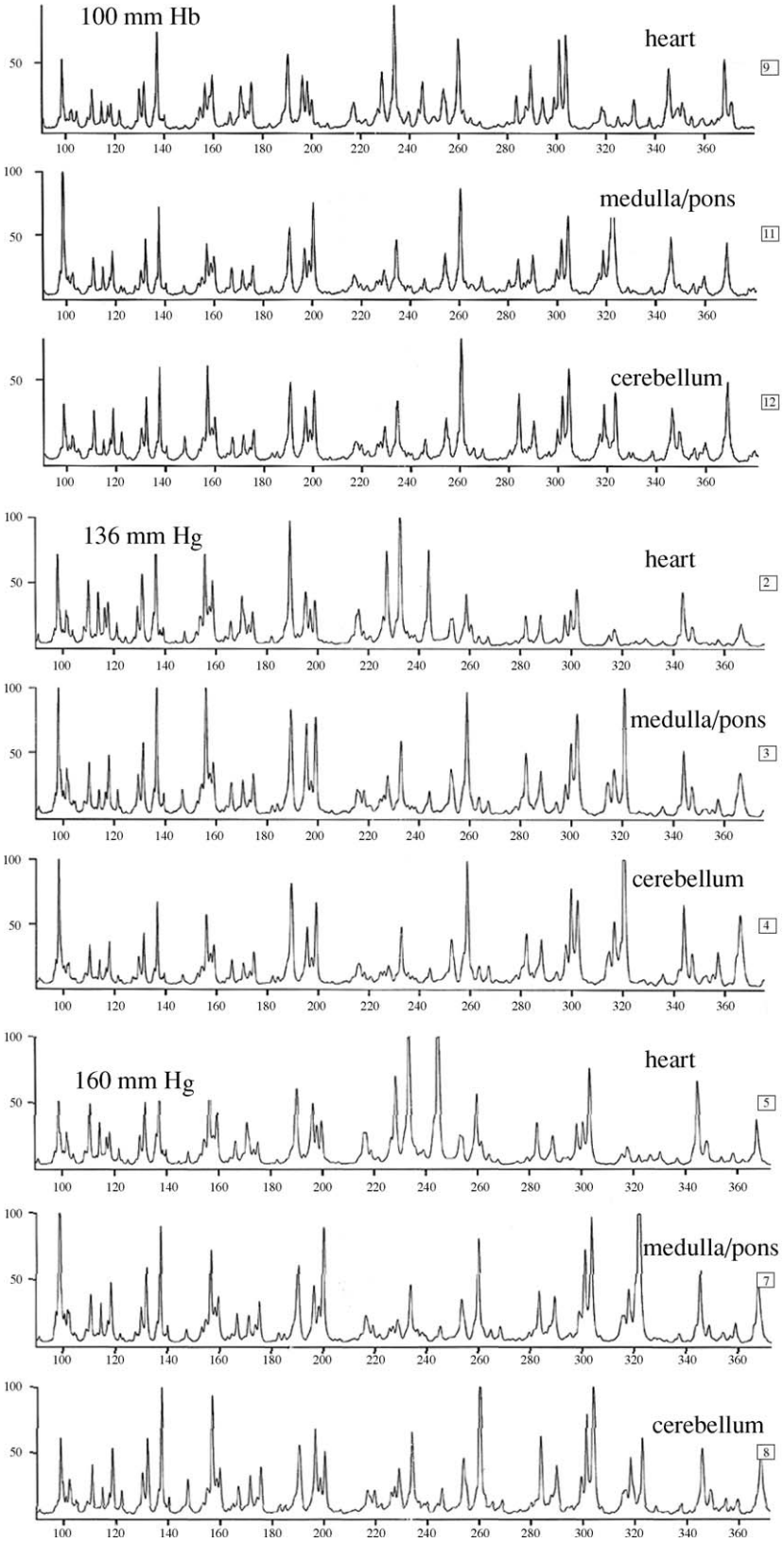

Fig. 6. Comparison of genes encoding zinc-finger mRNA expression levels in the heart, medulla/pons, and cerebellum of animal infused with NaCl. Conditions for this experiment are identical to those described in Fig. 3. Obvious tissues specific changes are visible in the 220–260 and the ~320 min ranges. Identification of other changes awaits quantitative analysis.

Table 3
Sequence analysis of random cDNAs from zinc-finger targeting experiments

Clone name	Sequences (5′–3′)[a]	bp[b]	Zinc-finger sequences[c]				Primer sequence[d]	%I[f]
			Rat		Mammal		5′ [CWGAA.KCAYA CDGGRGARAARCC] 3’[e]	
			%I d	*E*-value	%I d	*E*-value		
2CA1	GTGAGGTATGACATGCAAGAGAAGGCCTTCCCACATTCC ATACAATAAA	49	71	6.10E−05	82[h]	1.10E−04	**A**AGAA**T**CCATA **A**TGGGGAAAAACC	87
2CA2	ATAAGGTTGGAGCTATGGCTAAAGGYTTTACCACATTGG TTACACTTRTAGGGCTTCTCACCGTATGGACYCBCTCGTG	79	–	–	81[h]	9.70E−12	CAGAA**TCC**ACA **G**TGGA**A**AGAAACC	83
2CA3	AGATGCCGAGTGAGGTCTGACCTGCACTGAGAATGCC TTCCGTGCATCCCATGACATGTGAAAGGCTT CCCCCGTATGGGTCCTGT	86	84	1.10E−22	83[m]	5.40E−18	**ACTC**A.TCATACA GG**T**GAGAAGCC	83
2CC1	CACWTTBATCACACTTAAXVGGTTTCTCCCCTGTGTGCA CTCTCTXBSTGGATKCAGAGGCTGGMMCTBTGACTG AVGGCCTTTSCACACCCACAACACCTATRG	105	93	8.90E−25	93[h]	9.80E−13	**A**AGA**T**.TCACACT GGGGAGAAGCC	92
2CG1	GTGTATAGTKAGGTATGACMTGCAATAGAAGGCCTTCCC ACATTCCTTGACATTCATAG	59	–	–	83[h]	1.10E−06	**AG**GAACACA**C**A CAGGGGGAAAACC	74
7CA1	AGGGTATGACATGCAATAGAAGGCCTTCCCACAWTC MWTGACATYCATAG	50	84	8.70E−04	72[m]	0.003	**AC**GAA**T**CCACA CTGGGGAGAAGCC	87
7CA2	ATAAGGTWBGAGCTCTGCCTAAAGGCYTTACCACACT CAATTACACTTATAG	54	87	2.50E−07	85[h]	2.50E−07	**AG**GA**T.A**CACA CCGGAGAGAAGCC	79

Random PCR products were cloned and sequenced. BLAST queries did not contain the primer sequence.

[a] Sequence compared to Genbank using BLAST. Note that query sequence did not include the target primer sequence in order to avoid skewing BLAST results. The standard IACUC nomenclature is used for the degenerate nucleotide positions in the zinc-finger primers (see Table 2 legend).

[b] Length in bp of sequence used in BLAST search.

[c] Percent identity of highest score hitting a DNA sequence encoding a zinc-finger motif in rats and other mammal ([m]: mouse, [h]: human). *E*-value, like *p*-value, shows percentage of query to match the result by chance. The lower the percentage the more significant the hit is.

[d] Genomic sequence and percent identity to zinc-finger primer used as a target primer. Locating primer sequence in genomic sequence of BLAST results verifies the correctness of these hits.

[e] Zinc-finger primer sequence.

[f] Percent identity to primer sequence.

Table 4
Comparison of cDNAs targeting methods

Number	Description		Primer sequences (5′–3′)[a]		Success rate[b]	RNA source	Ref.[c]
	Base method	Target	Reverse transcription	PCR			
1	DD[d]	Prophenol -oxidase	T_{12}MA T_{12}MG T_{12}MC or T_{12}MT *[polyA tail primer with anchor bases]*	F: ACAAAGCTTCATRTGNTGNGGCCANCCRCANCCRCA R: ACAGAATTCATGMGNGAYCCNTTYTTYTAYMGNTGG *[Gene family primers with added EcoR1 and Hind III sites (underline)]*	10/11 (90%)	Fly	Hagen et al. (1997)
2	DD[d,e]	Coding sequences	T_{12-18} *[poly A tail primer]*	F: AGGCAGGA AGAAGGCC AGACCCTG AGAGGAGC ATGCTGGA CAAGGAGA CATCGTGG CCAAGAAG CCAGATGA CCTCACCA CCTGAAGG CCTGAGGA CCTGGAAA CCTGGAGA CTACCTGG CTGCAGAA CTGCCAAG GAAGCTGA GAAGTGGA GACCTGGA GAGCTGAA GAGGAGAT GGAGAAAG GTGGAGAA TCCTGGAC TGCTGAGC TGGACAAG TGGACCTG TGGAGAGA *[coding region of human genes]*	122/155 (78%) [computation test only]	Human	Lopez-Nieto and Nigam (1996)
		G-protein coupled receptors		CAACCCCATC CATCCTGGTG CATCTCCATC CCTGGCCATC CCTTCATCCT CTTCCTGCTG GACAGGTACA GCCTGGCCTG GCTATGCCAA GCTGCTGCTG GCTGGCTGCC GCTGGGCTAC GGACCGCTAC GGCCTTATC TCCTGSTGGC TCTGCTGGCT TCTGGGTGGC TGCTCATCAT TGCTGSTCAT TGGTGGCTGT TGGTGGCTGT *[G-protein coupled receptors]*	7/30 (23%)		

3	DD[d]	Zinc-finger	$T_{12}MG$ $T_{12}MA$ $T_{12}MT$ or $T_{12}MC$ *[poly A primer with anchor bases]*	F: GGNGAGAARCCCTWYGARTG R: Same primer used in the reverse transcription reaction	7/12 (58%)	Human	Johnson et al. (1996)
4	RAP[f]	Zinc-finger	GAGAGCTAGCTGAAMGTMG or GAGAGCTAGCTTATCWGGACC *[arbitrary primers]*	F: GTCGTCGAATTCCAYACHGGHGGRGAAAAACC GTCGTCGAATTCCAYACHGGHGGRGAAAAGCC GTCGTCGAATTCCAYACHGGHGGRGAGAAACC or GTCGTCGAATTCCAYACHGGHGGRGAGAAGCC *[Zn finger sequence for HC region]* R: Same arbitrary primer used in the reverse transcription reaction	5/12 (42%)	Mouse	Stone and Wharton (1994)
		Serine/ Threonine Protein kinase	CAGAGCTAGCTTNGCYTTRT *[arbitrary primer]*	F: GTCGTCGAATTCCGAYTTAGGGTT R: same arbitrary primer used in the reverse transcription reaction.	1/30 (3%)		
5	MI[g]	Zinc-finger	T_{18} *[but products further prepared by digestion with type IIS restriction enzymes and ligation to adapters]*	F: GTACATATTGTCGTTAGAACGC *[C1S primer]* or CGGAATTCGTACATATTGTCGTTAGAACGC *[CIT primer]* *[adapter primers]* R: CAYACIGGIAARAARCC *[zinc finger HC link primer]*	259 /1086 (24%)	Mouse	Yamashita et al. (2001)
6	DD[d]	*AU* motif	$G_7TATTTATTTANT_{15}Y$ $GGTGGGTGCTATTTATTTANT_{15}Y$ *[AU motif (TATAAT) = AU rich sequence found in 3' untranslated region]*	F: GGGGGGGTATTTATTTAA GGGIIIITATTTATTTA R: G_7TATTTATTAA G_7TATTTATTAC G_7TATTTATTAG G_7TATTTATTAT F: GGTGGGTGGTATTTATTA GGTIIITIITATTTATTA R: GGTGGGTGGTATTTATTTAA GGTGGGTGGTATTTATTTAC GGTGGGTGGTATTTATTTAG GGTGGGTGGTATTTATTTAT	8/59 (14%)	Human	Dominguez et al. (1998)
7	SPR[h]	Tyrosine Kinase	CCRAADGMCCABACRTCWGA CCRWARCTCCAVRCRTCRCT CCRAADGMCCABACRTCRCT and CCHADRCTCCAGACRTCRSA	F: AAATCGTAGTTATTGTCAGACGGTC *[adapter primer]* R: GTTSCGVGYRGCYAGGTCNCGRTG GTTWCKRGCWGCYAARTCYCGRTG RTTYCTDGCHGCCARGTCCCGRTG GTTTCKGGCAGCMAGRTYTCTATG RTTSKMRGGYTTCAMGTCTSTATG ATTRGARGGYTTSAYRTCTCTRTG *[conserved tyrosine kinase]*	~75	Human	Okada et al. (2002)

Table 4 (*Continued*)

Number	Description		Primer sequences (5′–3′)[a]		Success rate[b]	RNA source	Ref.[c]
	Base method	Target	Reverse transcription	PCR			
8	RAP[f]	Zinc-finger Serine/threonine protein kinase	Random hexamers	F: GGNGAGAARCCCTWYGARTG R: random hexamers F: CGGATCCAAMGNGAYYT R: random hexamers	One clone	Human	Donohue et al. (1995)
9	RAP[f]	Zinc-finger Leucine zipper	Random hexamers	F: GGNGAGAARCCCTWYGARTG R: CCHGTGTGARTCCTCTGRTG F: CTGGARGMNVAGRHSRMSMMGCT R: AGCTGGGYCRCCTTCTCCTCCAG	0%	Human	Hsu et al. (1993)
10	RAP[f,i]	Nuclear hormone receptor	A_{12}CATTTGGTC A_{12}GTCCTGGTC	F: GCTGCAAAAGITTCTT GCTGCAAGGTITTCTT R: same as reverse transcription primer	Not given	Rat	Yoshikawa et al. (1995)

[a] I: inosine. Standard IACUC nomenclature is used for the degenerate nucleotide descriptions (see Table 2 legend).

[b] The success rate is calculated as the number of clones with high (as defined by the authors) identity to known gene containing the target sequence/total number of sequences tested.

[c] Reference.

[d] DD (Liang and Pardee, 1992) – differential display method. This method uses a polyT primer with unique anchor bases along with a random primer. The random primer sequences have been optimized experimental.

[e] A set of octanucleotides (see below) that can detect ~75% human mRNAs was identified using simulated PCR with sense (shown below) and anti-sense (reverse complement of sequences shown below). The efficacy of a similar computer simulation analyzing G-protein coupled receptors was tested by cloning and sequencing the amplified, targeted, PCR products.

[f] RAP (Welsh et al., 1992) – RNA arbitrary primer method. RAP uses arbitrary primers only and no oligo dT primers.

[g] MI: molecular indexing (Kato, 1996). In this method, cDNAs made using an oligo$(dT)_{18}$ primers was digested by the IIS restriction enzymes *Fok*I and *Bsm*AI and ligated to adapters C1G (GTACATATTGTCGTTAGAACGCG and N[nnn]CGCGTTCTAACGACAATATGTAC) except when [nnn]*: AGA, GGA, TCC, ATC, CAG, CTG, TCT. In the later case, the digested cDNAs were ligated to CIT adapters (GTACATATTGTCGTTAGAACGCT and N[nnn]AGCGTTCTAACGACAATATCTAC. The type II restriction enzymes cleave outside of their recognition sequences leaving four base single-stranded overhangs. The pool of adapters with different 5′ end sequences ligated to different subsets of the cleaved fragments. [*[nnn] = 64 three base combinations].

[h] In this method, the reverse transcription primers were from the tyrosine kinase conserved domain. The cDNAs were digested with *Bst*YI or *Bsi*HKAI and then ligated to Y-shaped adapters. A double-stranded Y-shaped adapter has a complementary sequence adjacent to the ligation site and non-complementary adapter sequences (a1 and a2) distal to the ligation site (see below).

[i] The first round of PCR initiates from the target sequence (T) or its complement (T′). In the example given below, extension of a T primer annealed to a fragment with the sequence complementary to the target sequence (T′) will produce a fragment with a sequence complementary to a1 (a1′). The second round of PCR using the a1 primer sequence will amplify the product. Note, that different sequences will be amplified with different combination of T (vs. T′) and A1 (vs. A2) as the orientation of the target sequence may vary.

Most methods using one target primer are based on the original DD protocols and replace either the oligoT primer or the random primer with the target primer. Replacement of the random primer with a target primer only amplifies cDNAs with the target sequence near to the 3′ end however. Replacement of the oligoT primer with a target primer will amplify more mRNAs, as the location of the target sequence may be anywhere along the mRNA.

Two methods (numbers 5 and 7) described in Table 4 ligate adapter sequences to the ends of restriction enzyme cleaved cDNAs. Adapter primers are then used in the PCR amplification.

The level of targeting success for the studied sequences presented in Table 4 range up to 90% when two targeting primers were used on a cDNA library from flies. Several studies target zinc-finger encoding mRNAs at success levels that range up to 58%. This level is much lower that that encountered in this study.

It is difficult to determine the reason for a lack of targeting success in the published reports summarized in Table 4. Targeting success depends on the design of the primers and the PCR protocol. Some studies, including ours, designed primers for gene families before the genome of interest was sequenced.

The success rate of the TcDD approach described here is equally as high as the success rate of TGDD using total human genomic DNA as template (Bouchard et al., 1999; Lavrentieva et al., 1999; Broude et al., 1999). Therefore, we expect the cDNA targeting method described in this report to be successful for a wide range of selected sequences and organisms.

A primary goal of expression experiments is to identify different levels of gene expression. As with any gene expression protocol, each putative difference needs individual confirmation. The experiments reported here have not addressed that issue, but previous experiments with TGDD and preliminary experiments with TcDD (data not shown) indicate that quantitative cellular information is preserved at least to the extent of other, similar methods. Experiments to be described elsewhere (Smith et al., in preparation) have developed a quantitative method to compare the data generated in TGDD and TcDD experiments.

The results demonstrate the effectiveness of TcDD for analyzing gene families that contain simple repeat sequences as well as the more complex zinc-finger motif. In fact, the previously reported TGDD experiments are more difficult than the TcDD experiments reported here because the complexity of the starting DNA template complexitis larger.

In either TGDD or TcDD, the resulting pools of selected fragments can be analyzed in several manners. In this study, the fragments were analyzed by sequencing and by size fractionations. However, the selected fragments could also serve as input to DNA array experiments or sequencing experiments measuring gene expression. The advantage of using TcDD for DNA array experiments is that it pre-selects genes expressed at low levels before hybridization. This means that hybridization of such fragments would not be obscured by the higher signals generated from the faster hybridizing fragments present at higher concentrations. The sequencing experiments could analyze random members of the selected pool to determine relative levels, or use the selected fragments as input into SAGE or derivative experiments. In this case, the results would provide information on the relative amounts of a specific class of mRNAs.

Acknowledgement

This work was supported by grants NIH Score HL55001 and DOA AIBS2154 to CLS.

References

Alberts, B., 1994. Molecular Biology of the Cell. Garland Publisher, New York.
Ali, M., Markham, A.F., Isaacs, J.D., 2001. Application of differential display to immunological research. J. Immunol. Meth. 250, 29–43.
Becker, K.G., Nagle, J.W., Canning, R.D., Biddison, W.E., Ozato, K., Drew, P.D., 1995. Rapid isolation and characterization of 118 novel C_2H_2-type zinc finger cDNAs expressed in human brain. Hum. Mol. Genet. 4, 685–691.
Bonaldo, M.F., Lennon, G., Soares, M.B., 1996. Normalization and subtraction: two approaches to facilitate gene discovery. Genome Res. 6, 791–806.
Bouchard, J., Foulon, C., Strom, N., Nguyen, G.H., Smith, C.L., 1999. Analyzing genomic DNA discordance between monozygotic twins. In: Handbook of Molecular Genetic Techniques for Brain and Behavior Research: Techniques in the Behavioral and Neural Sciences, Elsevier, Amsterdam, The Netherlands, p. 994.
Bowtell, D., Sambrook, J., 2002. DNA microarrays: a molecular cloning manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Breslauer, K.J., Frank, R., Blocker, H., Marky, L.A., 1986. Predicting DNA duplex stability from the base sequence. Proc. Natl. Acad. Sci. U.S.A. 83 (11), 3746–3750.
Broude, N.E., Chandra, A., Smith, C.L., 1997. Differential display of genome subsets containing specific interspersed repeats. Proc. Natl. Acad. Sci. U.S.A. 94, 4548–4553.
Broude, N.E., Storm, N., Malpel, S., Graber, J.H., Lukyanov, S., Sverdlov, E., Smith, C.L., 1999. PCR based targeted genomic and cDNA differential display. Genet. Anal. 15, 51–63.
Chaudhari, N., Hahn, W.E., 1983. Genetic expression in the developing brain. Science 220, 924–928.
Cummings, C.J., Zoghbi, H.Y., 2000. Trinucleotide repeats: mechanisms and pathophysiology. Annu. Rev. Genomics Hum. Genet. 281–328.
Dominguez, O., Ashhab, Y., Sabater, L., Belloso, E., Caro, P., Pujol-Borrell, R., 1998. Cloning of ARE-containing genes by AU-motif-directed display. Genomics 54, 278–286.
Donohue, P.J., Alberts, G.F., Guo, Y., Winkles, J.A., 1995. Identification by targeted differential display of an immediate early gene encoding a putative serine/threonine kinase. J. Biol. Chem. 270, 10351–10357.
Hagen, H.E., Klager, S.L., McKerrow, J.H., Ham, P.J., 1997. *Simulium damnosum* s.l.: isolation and identification of prophenoloxidase following an infection with *Onchocerca* spp. using targeted differential display. Exp. Parasitol. 86, 213–218.
Hastie, N.D., Bishop, J.O., 1976. The expression of three abundance classes of messenger RNA in mouse tissues. Cell 9, 761–774.
Hatzinikolaou, P., Gavras, H., Brunner, H.R., Gavras, I., 1980. Sodium-induced elevation of blood pressure in the anephric state. Science 209, 935–936.
Hsu, D.K., Donohue, P.J., Alberts, G.F., Winkles, J.A., 1993. Fibroblast growth factor-1 induces phosphofructokinase, fatty acid synthase and Ca(2+)-ATPase mRNA expression in NIH 3T3 cells. Biochem. Biophys. Res. Commun. 197, 1483–1491.
Johnson, S.W., Lissy, N.A., Miller, P.D., Testa, J.R., Ozols, R.F., Hamilton, T.C., 1996. Identification of zinc finger mRNAs using domain-specific differential display. Anal. Biochem. 236, 348–352.
Kaplan, B.B., Schachter, B.S., Osterburg, H.H., de Vellis, J.S., Finch, C.E., 1978. Sequence complexity of polyadenylated RNA obtained from rat brain regions and cultured rat cells of neural origin. Biochemistry 17, 5516–5524.

Kato, K., 1996. RNA fingerprinting by molecular indexing. Nucl. Acids Res. 24 (2), 394–395.

Lavrentieva, I., Broude, N.E., Lebedev, Y., Gottesman, I.I., Lukyanov, S.A., Smith, C.L., Sverdlov, E.D., 1999. High polymorphism level of genomic sequences flanking insertion sites of human endogenous retroviral long terminal repeats. FEBS Lett. 443, 341–347.

Liang, P., Pardee, A.B., 1992. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. Science 257, 967–971.

Liang, P., Pardee, A.B., 1997. Differential display. A general protocol. Meth. Mol. Biol. 85, 3–11.

Lopez-Nieto, C.E., Nigam, S.K., 1996. Selective amplification of protein-coding regions of large sets of genes using statistically designed primer sets. Nat. Biotechnol. 14, 857–861.

Lorkowski, S., Cullen, P., 2003. Analyzing Gene Expression: A Handbook of Methods: Possibilities and Pitfalls, vol. 2. Wiley-VCH, Weinheim, NY.

Nguyen, G.H., Bouchard, J., Boselli, M.G., Tolstoi, L.G., Keith, L., Baldwin, C., Nguyen, N.C., Schultz, M., Herrera, V.L., Smith, C.L., 2003. DNA stability and schizophrenia in twins. Am. J. Med. Genet. B: Neuropsychiatr. Genet. 120, 1–10.

Okada, M., Yamaga, S., Yasuda, S., Weissman, S.M., Yasukochi, Y., 2002. Differential display of protein tyrosine kinase cDNAs from human fetal and adult brains. Biotechniques 32 pp. 856, 858, 860, 863–865.

Oliveira, R.P., Broude, N.E., Macedo, A.M., Cantor, C.R., Smith, C.L., Pena, S.D., 1998. Probing the genetic population structure of *Trypanosoma cruzi* with polymorphic microsatellites. Proc. Natl. Acad. Sci. U.S.A. 95, 3776–3780.

Ravasi, T., Huber, T., Zavolan, M., Forrest, A., Gaasterland, T., Grimmond, S., Hume, D.A., 2003. Systematic characterization of the zinc-finger-containing proteins in the mouse transcriptome. Genome Res. 13, 1430–1442.

Siebert, P.D., Chenchik, A., Kellogg, D.E., Lukyanov, K.A., Lukyanov, S.A., 1995. An improved PCR method for walking in uncloned genomic DNA. Nucl. Acids Res. 23 (6), 1087–1088.

Smith, C.L., Nguyen, G., Bouchard, J., Surdi, G., Yaar, R., Tolstoi, L., 2000. Comparative genomics: differential display and subtractive hybridization. In: Meyers, R.A. (Ed.), Encyclopedia of Analytical Chemistry: Instrumentation and Applications, vol. 6. John Wiley & Sons, Lmt., Sussex, UK, pp. 4893–4901.

Stallings, R.L., 1994. Distribution of trinucleotide microsatellites in different categories of mammalian genomic sequence: implications for human genetic diseases. Genomics 21, 116–121.

Stone, B., Wharton, W., 1994. Targeted RNA fingerprinting: the cloning of differentially-expressed cDNA fragments enriched for members of the zinc finger gene family. Nucl. Acids Res. 22, 2612–2618.

Sturtevant, J., 2000. Applications of differential-display reverse transcription-PCR to molecular pathogenesis and medical mycology. Clin. Microbiol. Rev. 13, 408–427.

Velculescu, V.E., Zhang, L., Vogelstein, B., Kinzler, K.W., 1995. Serial analysis of gene expression. Science 270, 484–487.

Vilain, C., Vassart, G., 2004. Small amplified RNA-SAGE. Meth. Mol. Biol. 258, 135–152.

Welsh, J., Chada, K., Dalal, S.S., Cheng, R., Ralph, D., McClelland, M., 1992. Arbitrarily primed PCR fingerprinting of RNA. Nucl. Acids Res. 20 (19), 4965–4970.

Yamashita, R., Matsubara, K., Kato, K., 2001. A comprehensive collection of mouse zinc finger motifs compiled by molecular indexing. Gene 274, 101–110.

Yoshikawa, T., Xing, G.Q., Detera-Wadleigh, S.D., 1995. Detection, simultaneous display and direct sequencing of multiple nuclear hormone receptor genes using bilaterally targeted RNA fingerprinting. Biochim. Biophys. Acta 1264, 63–71.

Young, B.D., Birnie, G.D., 1976. Complexity and specificity of polysomal poly(A+) RNA in mouse tissues. Biochemistry 15, 2823–2829.

Zhang, L., Zhou, W., Velculescu, V.E., Kern, S.E., Hruban, R.H., Hamilton, S.R., Vogelstein, B., Kinzler, K.W., 1997. Gene expression profiles in normal and cancer cells. Science 276, 1268–1272.