



ELSEVIER

## PCR based targeted genomic and cDNA differential display

Natalia E. Brode<sup>a,\*</sup>, Niels Storm<sup>a</sup>, Sarah Malpel<sup>a,b</sup>, Joel H. Graber<sup>a</sup>, Sergey Lukyanov<sup>c</sup>, Eugene Sverdlov<sup>c</sup>, Cassandra L. Smith<sup>a</sup>

<sup>a</sup> Center for Advanced Biotechnology and Departments of Biomedical Engineering, Biology and Pharmacology and Experimental Therapeutics, Boston University, Boston, MA 02215, USA

<sup>b</sup> Ecole Supérieure de Biotechnologie de Strasbourg, Boulevard Sebastien Brandt, Parc d'Innovation, Illkirch-Graffenstaden, Strasbourg, France

<sup>c</sup> Shemyakin-Ovchinnikov Institute for Bioorganic Chemistry, Russian Academy of Sciences, Moscow, Russia

Received 29 July 1998; accepted 10 November 1998

### Abstract

We previously described a targeted genomic differential display method (TGDD: Brode NE, Chandra A, Smith CL. Differential display of genomic subsets containing specific interspersed repeats. Proc. Natl. Acad. Sci. USA 1997;94:4548–53). In that method, presently characterized as method I, targeting was accomplished by capturing DNA fragments containing specific a sequence by hybridization with complementary single-stranded DNA. The captured fragments were amplified by PCR. Here, we describe method II where targeting is accomplished by PCR using primers specific to the target sequence. Method II takes advantage of PCR suppression to eliminate fragments not containing the target sequence (Siebert PDA, Chenchik A, Kellogg DE, Lukyanov KA and Lukyanov SA. An improved PCR method for walking in uncloned genomic DNA. Nucleic Acids Res 1995;23:1087–1088). Targeting focuses analysis on and around interesting areas and additionally serves to reduce the complexity of the amplified subset. These approaches are useful to amplify genome subsets containing a variety of targets including various conserved sequences coding for *cis*-acting elements or protein motifs. © 1999 Published by Elsevier Science B.V. All rights reserved.

**Keywords:** PCR based; Genomic; cDNA differential display

### 1. Introduction

Numerous methods are available to compare genomes. Many focus on comparing random fragments of genomic DNA. In simple organisms, this may involve a mere compilation of restriction fragment length polymorphism (RFLPs) or even DNA sequences directly. In contrast it is imperative that sample complexity be reduced in large genomes. Human DNA fingerprinting was accomplished by Southern hybridization with minisatellite sequences [1,2]. This methodology was largely replaced by a variety of polymerase chain reaction (PCR) methods. PCR methods that reduce complexity in a random fashion include randomly amplified fragment polymorphism (RAPD: [3–6]) and

amplified restriction fragment length polymorphism (AFLP: [7]). Some PCR fingerprinting methods use primers targeted to specific interspersed (e.g. inter-Alu PCR; [8,9]) sequences.

Simple repeat sequences have been used to assess genome stability during development and during disease processes such as cancer [10,11]. Simple repeat sequences, e.g. microsatellite sequences, are used extensively as genetic markers (e.g. sequence tagged sites (STSs); [12]) because they are extremely polymorphic. In most studies individual loci in different samples are tested and compared. Methods for isolating genomic DNA fragments containing interspersed simple repeat sequences are available. Most involve the screening of clone libraries for the repeat sequences. PCR-based approaches have been described that use [13–16] or do not use [12,17] a repeat capturing step before amplifica-

\* Corresponding author.

tion. For example, microsatellite sequences may be identified in cloned DNA samples (YAC's, P1, cosmids, bacteriophages or plasmid clones) by PCR using primers to a dinucleotide repeat and a vector sequence [18,19].

Subtractive hybridization is another way of comparing genomes. This method allows the isolation of an inserted or deleted sequence difference between two samples [20,21]. In complex genomes, the most successful subtractive hybridization experiments analyze random differences in expressed sequences (e.g. [22]). Differential display (DD: [23]) also analyzes random differences in expressed sequences. DD creates RNA fingerprints by displaying size fractionated randomly amplified cDNAs. The focus on RNA provides a simple and meaningful complexity reduction.

We recently described a comparative genomic method that focused analysis on and around  $(CAG)_n$  and  $(CGG)_n$  repeating sequences [24]. This method uses a sequence-specific capture step and PCR amplification to isolate genome subsets containing these sequences. Adapter-tagged genomic restriction fragments (e.g. restriction fragments ligated to known oligonucleotides) containing a target sequence are captured by hybridization to an immobilized complementary single stranded probe. The captured fragments are then amplified by PCR using only primers complementary to the adapter sequences (A-primer) or in the presence of primers complementary to the targeted sequence (T-primer). The PCR amplified and labeled fragments either are displayed sorted by size on a high resolution DNA sequencing gel as in DD to create a DNA fingerprint or are used to create libraries of cloned target sequences, e.g. of polymorphic simple repeats in *Trypanosoma cruzi* [25].

We also developed a second genomic targeting approach, solely using PCR, and used it to analyze genomic differences between monozygotic twins [26,27]. The second method uses a semi-nested PCR protocol and takes advantage of PCR suppression (PS: [28,22]) to amplify fragments containing the target and a flanking sequence. The details of method II are described in this manuscript.

## 2. Materials and methods

### 2.1. Materials

Nonphosphorylated oligonucleotides (Table 1) were purchased from Operon Technologies (Alameda, CA) and Cy5 labeled oligonucleotides were delivered from Amitof (Boston, MA). Genomic DNA was isolated from anonymous human blood and buccal smears. Monozygotic twin samples were obtained from cell lines provided by [29], and from a E. Fuller Torrey

Huntington Disease (HD) affected kindred provided by Jim Guzella (CEPH pedigree 102: [30]). Genomic DNA was prepared by standard phenol extraction procedure.

Human DNA from buccal smears was obtained by scraping both cheeks several times. Samples were suspended in 500  $\mu$ l of PBS buffer. Genomic DNA was isolated using a Qiagen Blood Kit according to the manufacturer's directions. Tissue samples from male Sprague-Dawleys rats for mRNA preparation were donated by Harry Gavras. mRNA was isolated from approximately 250 mg of heart, medulla/pons and cerebellum, using a Qiagen Oligotex Direct mRNA Kit (Qiagen, Hilden, Germany) with two final elutions of 50  $\mu$ l each.

### 2.2. Semi-nested PCR-based genomic targeting method

Phenol extracted genomic DNA (300 ng) was digested with 10 units of the restriction enzymes *Hae*III or *Sau*96 I at 37°C overnight in a 100  $\mu$ l total volume reaction. The recessed *Sau*96 I ends were filled in using 0.25 units of AmpliTaq (Perkin-Elmer, Norwalk, CT),

Table 1  
Oligonucleotides used

	Sequence (5' to 3')	Description
1	TGTAGCGTGAAGACGACA-GAAAGGGCGTG-GTGGCGAGGGCGGT	Adapter-40
2	ACCGCCCTCCG	Adapter-11
3	TGTAGCGTGAAGACGACAGAA	5' A-primer
4	AGGGCGTGGTGGC-GAGGGCGGT	3' A-primer
5	AGGGCGTGGTGGCAGGGCG-GTCCTT	<i>Hae</i> III 3' A-primer TT
6	AGGGCGTGGTGGCAGGGCG-GTCCTG	<i>Hae</i> III 3' A-primer TG
7	AGGGCGTGGTGGCAGGGCG-GTCCTT	<i>Hae</i> III 3' A-primer TTT
8	AGGGCGTGGTGGCAGGGCG-GTCCTC	<i>Hae</i> III 3' A-primer TTC
9	AGGGCGTGGTGGCAGGGCG-GTCCTA	<i>Hae</i> III 3' A-primer TTA
10	AGGGCGTGGTGGCAGGGCG-GTCCTG	<i>Hae</i> III 3' A-primer TTG
11	AGGGCGTGGTGGCAGGGCG-GTCCTC	<i>Hae</i> III 3' A-primer TTCG
12	A(A/C/T)(CTG) <sub>6</sub>	3' T-primer A(A/C/T)
13	(CTG) <sub>6</sub> A	5' T-primer A
14	(CTG) <sub>6</sub> AC	5' T-primer AC
15	(CTG) <sub>6</sub> AT	5' T-primer AT
16	AGGGCGTGGTGGCAGGGCG-GTGCCCA	<i>Sau</i> 96 I 3'-A-primer GA
17	ATGAAGGCCTTCGAGTCCCT-CAAGTCCTT	primer HD1F
18	GGCGGCTGAGGAAGCT-GAGGA	primer HD1R
19	CGCAGTCGACCGT <sub>13</sub> (A/G/C)	First strand synthesis primer

in Perkin-Elmer's 1X PCR buffer I (10 mM Tris-HCl, pH 8.3, 50 mM KCl, 1.5 mM MgCl<sub>2</sub>, 0.001% (w/v) gelatin) and 100 μM dNTPs at 72°C for 1 min. DNA was ethanol precipitated, dissolved in 20 μl of sterile water and blunt-end ligated to an excess (2 μM each) of complementary adapter oligonucleotides (numbers 1 and 2, Table 1) at 16°C overnight in a 30 μl final volume, containing 50 mM Tris-HCl, pH 7.6, 10 mM MgCl<sub>2</sub>, 0.5 mM ATP, 10 mM dithiothreitol and 5 units of T4 DNA ligase (Life Technologies, Gaithersburg, MD). The complementary oligonucleotides used as adapters were of different lengths to insure the adapters ligated to the blunt ended genomic fragments with the same polarity. The ligation reactions produced genomic restriction fragments with 5' 26-base single-stranded overhangs. The ligation was terminated by incubation at 75°C for 5 min. The adapter-tagged restriction fragments were purified from unligated adapter oligonucleotides by passing the samples through Wizard DNA purification columns (Promega, Madison, WI) and eluted into 50 μl of sterile water.

Usually, the prepared DNA fragments (3–5 ng) were amplified by PCR in a 50 μl reaction volume in Perkin-Elmer's PCR buffer II (10 mM Tris-HCl, pH 8.3, 50 mM KCl) plus 2.5 mM MgCl<sub>2</sub>, 250 μM of each dNTP and 2.5 units of AmpliTaq DNA polymerase. Hot start PCR was performed at 94°C by adding 0.2 μM each of the adapter-primer (A-primer) and fluorescein labeled CTG-containing target-primer (T-primer), e.g. oligonucleotides 3 and 12 or 14 (Table 1), respectively, or by the addition of AmpliTaq. Variations from this basic protocol are indicated in the results section.

The PCR mixtures were subjected to 20–25 amplification cycles consisting of incubations at 94°C for 3 s, 65°C for 20 s and 72°C for 30 s in the PTC-100™ Temperature Cycler (MJ Res., Watertown, MA). The products of this first PCR reaction were diluted 1000-fold and used as templates for a second PCR amplification. In the second PCR amplification oligonucleotide 3 in the first reaction was replaced by one of seven oligonucleotides 4 through 11 (Table 1). The anchor bases vary in A-primers designed 4–11 in Table 1. Anchor sequences anneal the primer to genomic sequences adjacent to the *Hae*III site and are used to reduce the complexity of the PCR product.

The PCR products (1–2 μl) were denatured for 3 min at 90°C in a stop solution (Pharmacia Biotech, Upsala, Sweden) containing 6 mg/ml of dextran blue and 0.1% sodium dodecyl sulfate in deionized formamide, loaded onto a 6% denaturing polyacrylamide gel (PAAG) and analyzed on the ALF DNA sequencing instrument (Pharmacia Biotech). The results were visualized using the Fragment Manager software provided with the instrument. A fluorescein-labeled 50 base pair (bp) ladder (50–500 bp; Pharmacia Biotech) was used as a size marker.

### 2.3. TGDD of HD alleles

TGDD of *Sau*96 I restriction fragments from DNA of HD-affected individuals was executed as described above, except that oligonucleotides 12 and 16 (Table 1) were used in the second PCR. The PCR products obtained were diluted 500-fold and reamplified by PCR with oligonucleotide 12 and HD-specific oligonucleotide 17 (Table 1). To estimate the length of CAG repeats in the Huntington gene a HD-specific PCR was performed as described in [24] using HD-specific primers (oligonucleotide 17 and 18, Table 1). Amplified products were analyzed on an ALF DNA sequencer after ten cycles.

### 2.4. Targeted DD (TDD) of cDNAs containing (CAG)<sub>n</sub>-repeat sequences

cDNA synthesis was performed using the Clontech PCR-Select™ cDNA Subtraction Kit (Clontech, Palo Alto, CA) as recommended by the manufacturer. Isolated mRNA (~2 μg) was ethanol precipitated, redissolved in 4 μl of sterile water, mixed with 20 pmoles (1 μl) of first strand primer (oligonucleotide 19, Table 1), heated to 70°C for 2 min and then cooled on ice for 2 min. The first strand was synthesized in a 10 μl reaction containing 50 mM Tris-HCl (pH 8.5), 8 mM MgCl<sub>2</sub>, 30 mM KCl, 1 mM dithiothreitol, 1 mM each dNTP and 20 units of AMV reverse transcriptase. After incubation for 90 min at 42°C, the reaction tubes were kept on ice while the following components were added: 48.4 μl sterile water, 16 μl 5X second strand synthesis buffer (1X = 100 mM KCl, 10 mM NH<sub>4</sub>SO<sub>4</sub>, 5 mM MgCl<sub>2</sub>, 0.15 mM β-NAD, 20 mM Tris-HCl (pH 7.5), 50 μg/ml bovine serum albumin), 1.6 μl dNTP mixture containing 20 pmoles of each deoxynucleotide triphosphate and 4 μl of an enzyme cocktail containing 24 units DNA polymerase I, 1 unit RNase H and 4.8 units of *Escherichia coli* DNA ligase. The reaction was incubated for 120 min at 16°C, 2 μl (6 units) of T4 DNA polymerase were added and the sample was incubated for an additional 30 min at 16°C. The reaction was stopped by addition of 10 mM EDTA-50 μg/ml glycogen (final concentrations). The double stranded cDNAs were purified by phenol/chloroform/isoamyl alcohol extraction, ethanol precipitated and redissolved in 50 μl of water. Samples (600 ng) were digested with 20 units of restriction enzyme *Rsa*I and ligated to adapter oligonucleotides (oligonucleotides 1 and 2, Table 1) as described above. Excess adapters and enzyme were removed using a Qiaquick PCR purification Kit (Qia-gen, Hilden, Germany).

A semi-nested PCR was carried out to target (CAG)<sub>n</sub> fragments. cDNA (5–10 ng) was first amplified in a 25 μl PCR (50 mM Tris-HCl, pH 8.3, 16 mM NH<sub>4</sub>SO<sub>4</sub>, 5 mM MgCl<sub>2</sub>, 150 μg/ml bovine serum albumin and 200

$\mu\text{M}$  each dNTP) with 0.2  $\mu\text{M}$  of oligonucleotide 3 and Cy5 labeled target oligonucleotide 13 (Table 1) and 9.4 units (or 1.25 units in later experiments) of KlenTaq1 DNA polymerase (Ab Peptides, St. Louis, MO) in 22 or 25 cycles (94°C for 15 s, 68°C for 30 s and 72°C for 60 s). Two  $\mu\text{l}$  of 1:100 diluted first PCR were used as template in a second PCR (25  $\mu\text{l}$  reaction volume with 10 mM Tris-HCl, pH 8.3, 50 mM KCl, 5 mM MgCl<sub>2</sub> and 200  $\mu\text{M}$  of each dNTP) with 0.2  $\mu\text{M}$  of oligonucleotide 6 (Table 1), 0.2  $\mu\text{M}$  of the same Cy5 labeled T-primer used in the first PCR and 1.25 U AmpliTaq polymerase (Perkin-Elmer, Norwalk, CT) in 20 or 22 cycles (cycle conditions as described above). All PCRs were performed using TaqStart antibody (Clontech) as directed by the manufacturer to guarantee hot start.

### 2.5. Cloning and DNA sequencing of targeted amplification products

Randomly chosen PCR amplification products obtained from *Sau96* I-digested DNA were cloned and sequenced to investigate the specificity of the method. In these experiments, oligonucleotides 12 and 3 (first PCR; Table 1) and oligonucleotides 12 and 16 (second PCR; Table 1) were used as the T- and A-primers, respectively. The PCR products from the second PCR amplification were cloned using a TA Cloning Kit (Invitrogen, Carlsbad, CA). Plasmid DNAs were isolated and sequenced using a Sequenase 2.0 Kit (Pharmacia Biotech) and an ALF sequencing instrument.

To isolate differential fragments specific DNA bands were excised from a PAAG (see above) or 5% 3:1 (weight:weight) NuSieve: LE agarose (obtained from FMC BioProducts, Rockland, ME), purified using the GeneClean Kit (Bio 101, Vista, CA), reamplified by PCR using the same oligonucleotides used in the second PCR of the original TGDD experiment and cloned using the TA Cloning Kit from Invitrogen; 4–6 clones were sequenced.

## 3. Results

### 3.1. Principles of a semi-nested PCR based TGDD method

The two-step semi-nested PCR-based amplification scheme is shown in Fig. 1. Briefly, the PCR templates are tagged genomic restriction fragments. The tags are 40-base pair (bp) oligonucleotides (adapters) ligated to the ends of the restriction fragments. A GC-rich 40-base adapter sequence is used to promote the annealing of complementary adapter sequences to each other, forming a hairpin structure. The annealed adapter ends suppress annealing/extension of shorter PCR primers (A-primers) complementary to the adapter sequence.

This effect has been called PS for PCR suppression [28,22].

Under PS conditions, efficient PCR amplification is achieved when two primers are used. One 20-mer A-primer corresponds to one of the self-complementary end adapter sequences. The second oligonucleotide, termed the T-primer, is complementary to a target sequence located in the single-stranded section of end-annealed genomic fragments. Single-stranded PCR products, produced by extension of an annealed T-primer no longer have complementary ends. Hence, these fragments will not be subjected to PS and will be efficiently PCR amplified. Meanwhile, genomic fragments that do not contain the target sequence remain end-annealed and are not replicated using the A-primer. Occasional extension of an annealed A-primer to an original template containing two adapter end sequences will produce a single-strand fragment that is subjected to PS because of its complementary end sequences. Thus, PS enhances targeting by inhibiting PCR amplification of fragments not containing the target sequence.

Most of the experiments described here target CAG-repeat sequences. It should be noted that the use of a T-primer containing a CTG-repeat amplifies a unique sequence flanking one side of the repeat sequence, whereas the other flanking sequence is amplified when the T-primer contains a CAG-repeat sequence.

### 3.2. Anchored PCR controls the complexity and composition of the selected PCR products

The complexity of a particular genome subset analyzed by TGDD is modulated by the total number of occurrences of the targeted sequence and the distance to adjacent occurrences of the restriction enzyme recognition site. The size range of the method used for analysis is also important for complexity control. The use of anchored PCR primers aids targeting by allowing a further controlled complexity reduction. Anchored primers have unique nucleotides at their 3' end or at their 5' end. The 3' unique nucleotides are used to anchor the primer to a unique genomic sequence adjacent to the target or the adapter sequences so that only a subset of the fragments is amplified.

Targeted genomic differential display (TGDD) of (CAG)<sub>n</sub>-repeat containing genomic *Hae*III fragments is shown in Fig. 2. In these experiments *Hae*III digested genomic DNA was ligated to adapters designated oligonucleotides 1 and 2 in Table 1. In the first PCR amplification, the A-primer was a 21-mer (oligonucleotide 3, Table 1) corresponding to the outermost portion of the ligated adapter oligonucleotide. The fluorescein-labeled T-primer (oligonucleotide 12, Table 1) was composed of a 3' CTG-repeating sequence plus two unique anchor nucleotides at the 5' end.

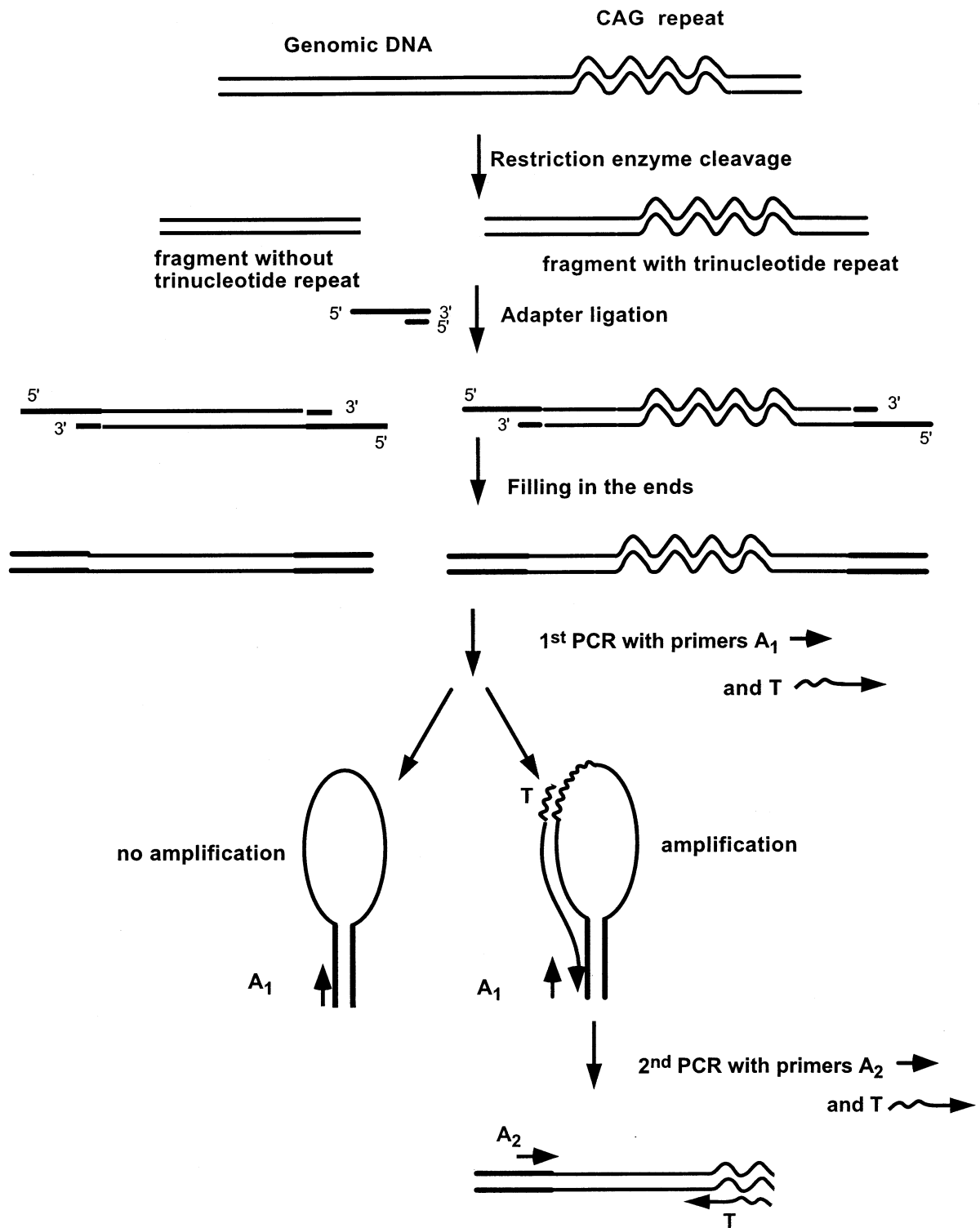


Fig. 1. Schematic of the semi-nested PCR protocol developed for TGDD. The first PCR uses a T-primer complementary to the target sequence and an A-primer complementary to the outermost portion of the adapter sequences ligated onto the end of the genomic restriction fragments. The second PCR uses the same T primer and an A-primer complementary to the 3'-portion of the ligated adapter sequence (see text for details).

The products of the first PCR were diluted and used as templates in a second PCR. The first and second PCR used the same T-primers but different A-primers. The A-primer in the second PCR reaction was a 26- or

27-mer oligonucleotide containing a 5' terminal 22-base segment complementary to the innermost adapter sequence plus a 3' terminal tetranucleotide (CCTT) or pentanucleotide (CCTTA, CCTTG, or CCTTT) se-

quence. The CC dinucleotide anneals to *Hae*III sequences remaining on the restriction fragments. The 3' terminal TT, TTA, TTG and TTT bases (lanes 1, 2, 3 and 4, respectively, Fig. 2) anchored the A-primers to complementary genomic sequences adjacent to the *Hae*III recognition site. Complexity reduction by anchor PCR is roughly proportional to the length of the anchor (e.g. dinucleotide and trinucleotide anchors should reduce the CAG-containing *Hae*III fragment

complexity by 16-fold and 64-fold, respectively) and more specifically proportional to the frequency of occurrence of the anchor sequence. For instance, the relative frequency of occurrence of a 3' CA dinucleotide is 34% compared to 5–10% for the other possible dinucleotide anchors in primate sequence from GenBank release 106.0 on 15 May 1998.

The data in Fig. 2 shows that different genome subsets were displayed when the same T-primer was

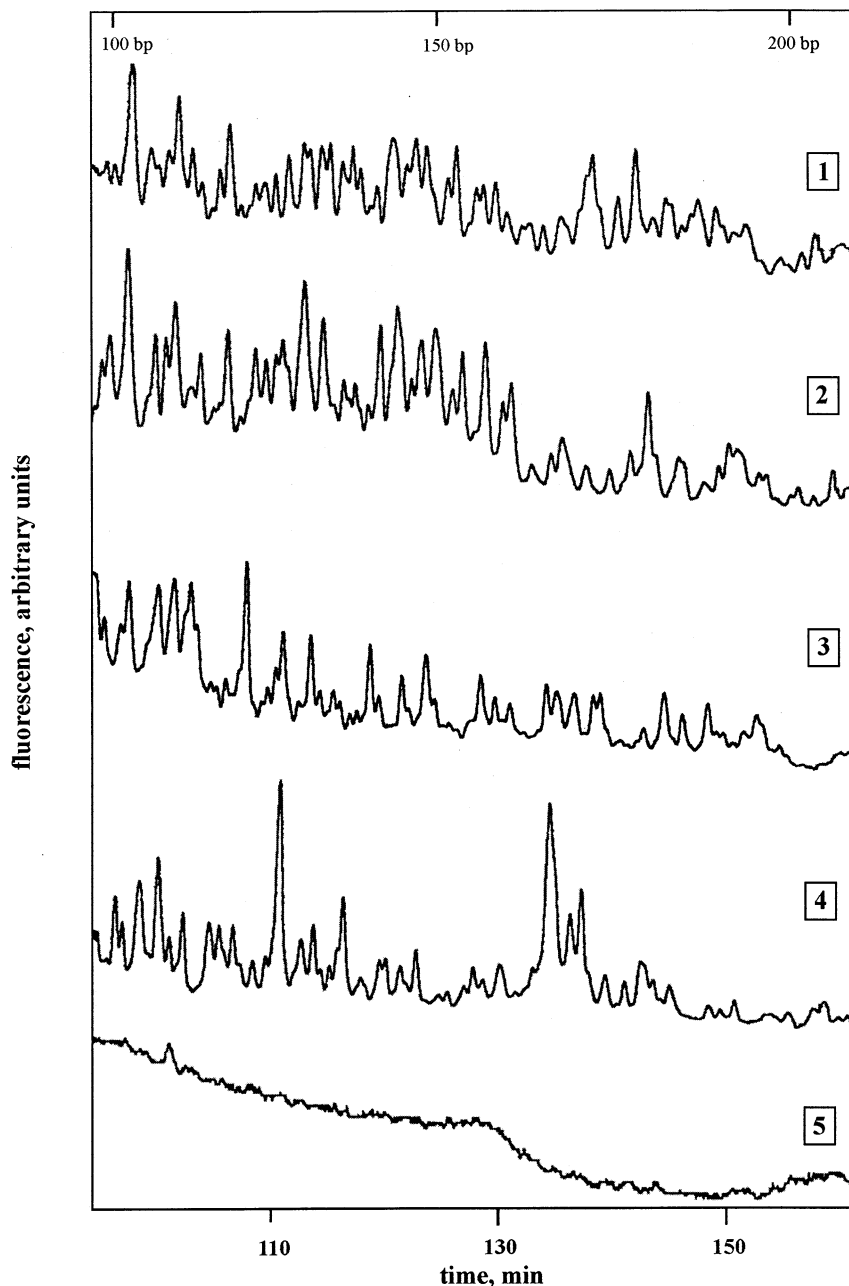


Fig. 2. TDGG of  $(CAG)_n$ -containing *Hae*III fragments. DNA samples digested were amplified first with a fluorescent labeled T-primer and an A-primer (oligonucleotides 12 and 3, respectively, Table 1) and then with the same T-primer and various A-primers (oligonucleotides 5 (lane 1), 10 (lane 2), 9 (lane 3), or 7 (lane 4), Table 1) as described in the text. No PCR products were generated when the fluorescent labeled T-primer was used alone (lane 5). Autoscaled fluorescence intensity versus time of elution is shown. The data in each lane was independently autoscaled by Fragment Manager software provided by the ALF manufacturer.

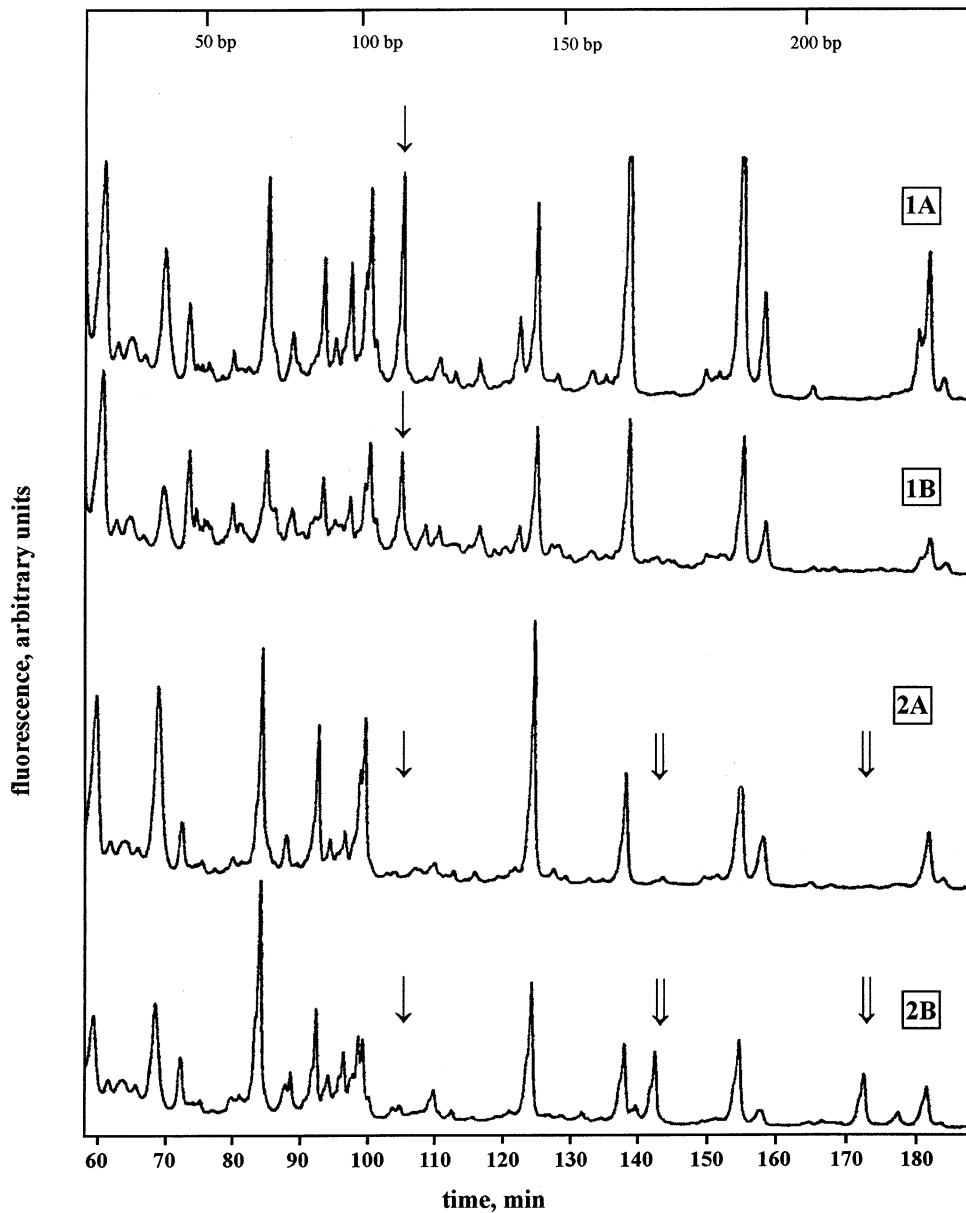


Fig. 3. TGDD comparison of  $(CAG)_n$ -repeat containing genomic subsets from two pairs of monozygotic twins (pair # 1 lanes 1 and 1A; pair # 2 lanes 2 and 2A). Both PCR amplifications used a 3'-anchored fluorescein labeled T-primer (oligonucleotide 12, Table 1). The A-primers were oligonucleotides 3 and 8 (Table 1) in the first and second PCR amplifications, respectively. Arrows indicate differences between unrelated twin pairs ( $\downarrow$ ) and between individuals of one twin pair ( $\Downarrow$ ). See text and Fig. 2 legend for details.

used in combination with A-primers having the same adapter sequence but varying anchor sequences. PCR amplification depended on the presence of both an A- and a T-primer, since no PCR products were detected when these oligonucleotides were used alone (e.g. lane 5). PCR using an A-primer with a 3 base anchor and a 2 base 5' anchored T-primer were used in the experiment shown in Fig. 2. Several hundred fragments within the size range of  $\sim 100$  to  $\sim 800$  bp (lane 2–4, Fig. 2) were amplified.

A complexity reduction was achieved using 3' terminal anchors on both A- and T-primers (Fig. 3). These

experiments used DNA samples from monozygotic twins. PCR amplification was performed with a 3'-terminal anchored T-primer with two anchor bases (oligonucleotide 14, Table 1) and a 3'-terminal anchored A-primer with three anchor nucleotides (oligonucleotide 8, Table 1). In these experiments the number of detected peaks was reduced from several hundred (compare Fig. 2) to 9 distinguished clearly visible peaks in the size range between 100 and 400 bp (Fig. 3 shows only a part of the display). This reduction correlates with the theoretically expected complexity reduction from  $4^3$  (64-fold) to a  $4^5$  (1024-fold) reduction when 3

and 5 total anchor bases, respectively, are used. Although a general conservation in the pattern of the displayed fragments was observed, clear differences between the twin pairs were also noted when an identical primer set was used for amplification.

A T-primer composed solely of a tandemly repeated sequence may anneal to any location within the target tandem repeat. Furthermore, the length of the repeat will be successively shortened with each round of amplification until it reaches a minimum equal to the length of the tandem repeat in the T-primer itself. Anchor nucleotides in primers targeting tandemly re-

peated sequences function to reduce complexity but also to target the primer to the 3' or 5' end of the tandem repeat. The data in Fig. 3 also show the differences in DNAs of monozygotic twins (compare lanes 3A and 3B, the differences are marked by arrows).

### 3.3. Reproducibility of TGDD

Reproducibility was assessed by comparing displayed pattern obtained using DNA isolated from the same individual several times (Fig. 4). In the depicted experiment, TGDD was accomplished with (CAG)<sub>n</sub>-contain-

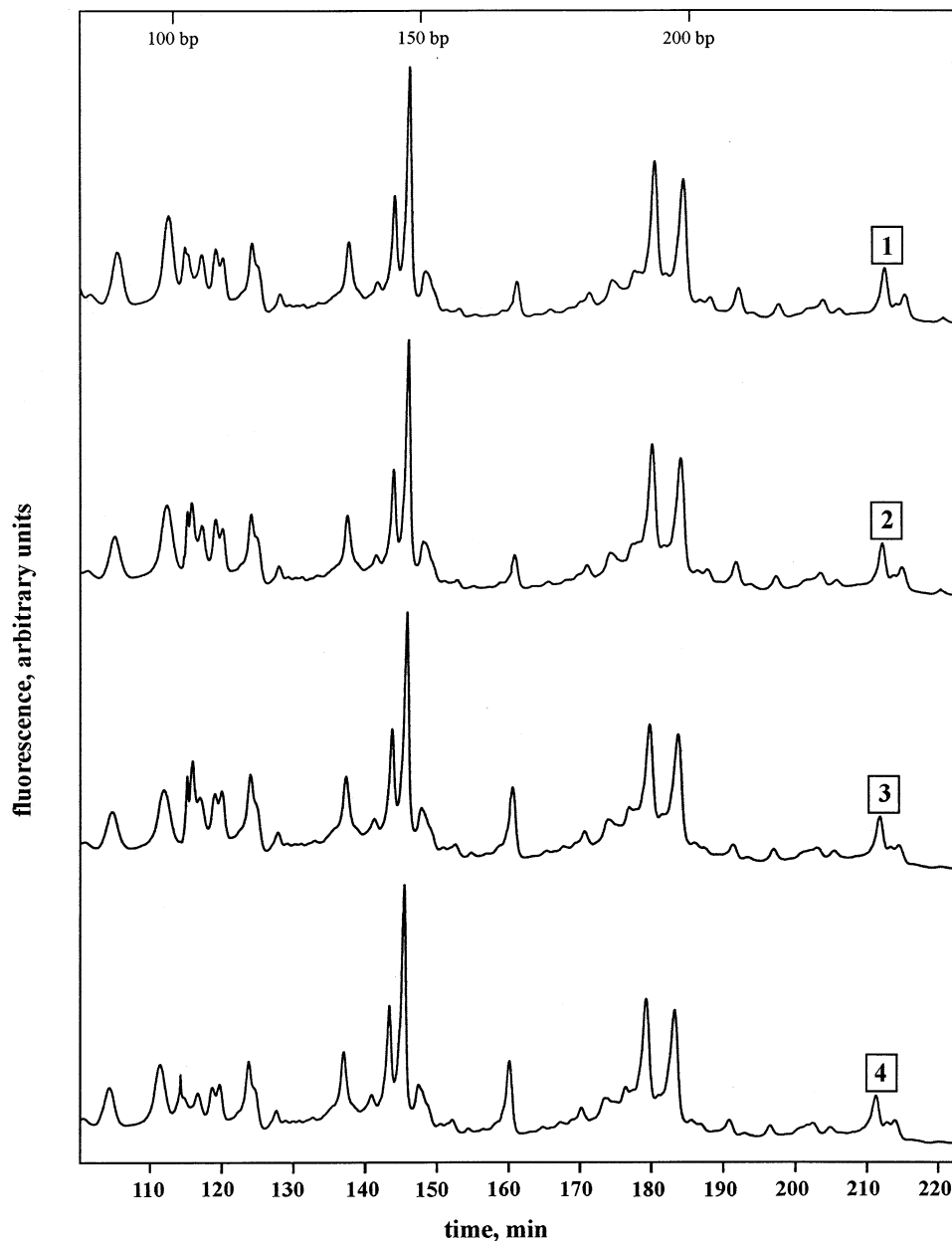


Fig. 4. Reproducibility of TGDD targeting of (CAG)<sub>n</sub>-containing sequences. Genomic DNA from the same individual was isolated twice from blood (lanes 1 and 2) and twice from buccal smears (lanes 3 and 4). TGDD was done using A- and T-primer oligonucleotide 14 and 9, respectively. See text and Fig. 2 legend for details.



ing genomic fragments obtained from genomic DNA templates isolated several times from the same blood sample (lane 1 and 2) or from buccal scrape of the same individual (lane 3 and 4). PCR amplifications were utilized with a 5' terminally anchored T-primer (oligonucleotide 12, Table 1) in combination with a nested *Hae*III A-primers. The results show that the same pattern of fragments is displayed when the same DNA is isolated multiple times from the same tissue or from different tissues of the same individual.

### 3.4. Effectiveness of targeting

The sequences of ten randomly chosen cloned products were determined from a *Sau*96 I genome subset targeted to CAG-repeat containing sequences (Table 2). In this experiment, the first PCR reaction used oligonucleotides 3 and 12, whereas the second PCR reaction used oligonucleotides 16 and 12 (Table 1) as the A-primer and fluorescein-labeled 5' anchored T-primer, respectively. All of the clones contained 6–8 perfectly, or imperfectly, repeated CAG trinucleotides at one terminus. Clones 3, 4, 5, 8 and 10 also contained multiple scattered CAG repeats nearby the tandem repeat. Scattered CAG repeats are commonly found near tandem CAG repeats [31]. Six clones contained the exact repeat primer sequence and four lacked the 5'-distal anchoring base. The sequences shown in Table 2, minus the primer sequences, were filtered and compared to DNA sequences in GenBank using BLAST. Filtering removes tandem repeating sequences from the analysis to avoid incorrect alignments based on low complexity sequences. Clones 3 and 4 matched previously identified sequences containing CAG-repeats [31]. Clone 10 contained a gene fragment with a *Sau*96 I restriction site upstream of the CAG-repeat sequence [32]. These results demonstrate high specificity of targeting, more than 90% of clones contain the targeted sequence. Note that similar cDNA studies obtained only 10–20% of products with the target sequence [33].

The prevalence of short repeats lengths in the cloned sequences correlates with the predominance of short CAG repeats (5–7 repeats) in the genome [34–36]. However, the results are also consistent with the preferential PCR amplification of short repeat lengths [38,39,24]. Increasingly stable hairpin structures formed from CAG-or CTG-repeating sequences of increasing length (for review see [37]) appear to inhibit PCR amplification. This means that short repeat sequences are preferentially amplified [38,39,24] and are distinguishable from long repeat sequences [38,24].

### 3.5. TGDD of HD alleles

A series of experiments, similar to those described by us [24], were performed to demonstrate that different

Table 2

DNA sequences of randomly chosen cloned fragments obtained after TGDD using a semi-nested PCR protocol

Clone #	Length of insert (bp)	DNA sequence <sup>a</sup>
1	133	AT(CTG) <sub>7</sub> - TGGAACAAACTCCAG- GATTTAGC...
2	153	AT(CTG) <sub>6</sub> CAATTATTATGAA...
3 <sup>b,c</sup>	209	ACTGTTGCTGTGGAACTG- TAGCTGTTGCTG- TGAAAACCTGATGCG- GATGGGCTTG(CTG) <sub>5</sub> - TTGCTG...
4 <sup>d</sup>	255	AA(CTG) <sub>6</sub> CCAG(CTG) <sub>2</sub> - GGTCAGGATAGGAGAG- GCTG...
5	193	A(CTG) <sub>6</sub> - CCATAACTGCTTCCACTC- CAAAGGC...
6	147	AA(CTG) <sub>6</sub> - TGGAACAAACTCCAG- GATC...
7	120	C(CTG) <sub>6</sub> - GCCTGAAGACCACATTCTGA- GAACCCCTGA...
8	216	AA(CTG) <sub>6</sub> GTGGGG(CTG) <sub>9</sub> - GTGCTGGTGGTGGGG(CTG) <sub>7</sub> ...
9	126	AT(CTG) <sub>6</sub> - GCCTGACTACTGACCACA...
10 <sup>e</sup>	102	A(CTG) <sub>6</sub> - TGTTGGCTGTGCTGGTTGTT- GCTG...

<sup>a</sup> Only the terminal sequence including the repeat sequence in the T-primer is shown. GenBank accession numbers for the entire sequences for clones 1–10 are U92822–U92824 and U92826–U92832. Sequence 10 is a fragment of transcriptional activator hSNF2a gene [32].

<sup>b</sup> Clone 3 did not contain repeat sequence of the T-primer.

<sup>c</sup> BLAST homology score is 4xe-28 to accession number X76572.

<sup>d</sup> BLAST homology score is 5.6xe-72 to accession number X73969.

<sup>e</sup> BLAST homology score is 8.4xe-27 to accession number D26155.

HD alleles could be detected by the PCR-based TGDD method described here. These experiments were required because the unmatched samples with different genetic background (three from an HD-affected kinship, HD-A, HD-B and HD-C and one anonymous control) used in these experiments were expected to and did display, many differences. Hence, a number of experiments were needed to confirm the identity of HD-containing fragments in the displays. Previously

performed experiments, using HD-specific PCR amplification had established that the HD-A and the control samples had two normal length  $\sim 20$  (CAG)<sub>n</sub>-repeats, whereas the HD-B and HD-C samples were homozygous expanded for  $\sim 40$ –46 repeats (Fig. 6A), as shown by [24]. TGDD of *Sau96* I fragments containing (CAG)<sub>n</sub> subsets is shown in Fig. 6B. The arrow indicates the location of the fragment expected to have the normal length allele. The identity of the HD containing fragment was confirmed by subjecting the samples shown in Fig. 6B to an HD-specific PCR using one HD-specific primer along with a T-primer (Fig. 6c).

A heterogeneous  $\sim 100$ –110 bp product was present in the HD-A and control and absent in the HD-B and HD-C samples. The expected size of a normal length HD-allele was 110 bp when the 5'-anchored primer annealed to the 5' end of the (CAG)<sub>n</sub> repeat. These results corroborate recent experiments by ourselves [24] and others [3,38] on biased amplification of short repeat lengths.

### 3.6. De novo identification of differences between samples

TGDD was developed to target simple repeat sequences. Since the T-primer can be anchored with unique nucleotides at its 3' or 5' end, sample comparisons can be focused solely on the unique sequences flanking the targeted sequence or on the sequences containing part of the repeat and flanking unique sequences together, respectively (see above).

Methylation or base changes within the restriction enzyme recognition site have the potential to be detected along with primer mismatches. One randomly chosen polymorphism was characterized in more detail (see Fig. 3). The identified fragment was about  $\sim 100$  bp in length. The fragment was isolated, cloned and sequenced as described in Section 2. A set of nested PCR primers from the unique sequence adjacent to the target (CAG)<sub>n</sub> sequence was used along the nested A-primer reaction to amplify the polymorphic target in two DNA samples, 1B and 2B. Six and four, randomly chosen clones were sequenced from the samples 1B and 2B, respectively in Fig. 3. The sequencing results (Fig.

	1			40
DNA 1B	TTTGCAATAA	<u>ATCCTGCTGC</u>	<b>TGCTCACTCT</b>	TTGGATCCAC
DNA 1B'	TTTGCAATAA	<u>ATCTTGCTGC</u>	<b>CTCACTCT</b>	TTGGGTCCAC
DNA 1B''	TTTGCAATAA	<u>ATCTTGCTAC</u>	<b>TGCTCACTCT</b>	TTGGGTCCAC
DNA 2B	TTTGCAATAA	<u>ATCTTGCTGC</u>	<b>TGCTCACTCT</b>	TTGGGTCCAC
DNA 2B'	TTTGCAATAA	<u>ATCTTGCTGC</u>	<b>TGCTCACTCT</b>	TTGGGTCCAC
DNA 2B''	TTTGCAATAC	<u>ATCTTGCTAC</u>	<b>TGCTCACTCT</b>	TTGGGTCCAC

Fig. 5. DNA sequence analysis of polymorphic fragments revealed by TGDD targeting of (CAG)<sub>n</sub>-containing restriction fragments (see Fig. 3, sample 1B and 2B). Differences are marked in bold and the location of the primer annealing site is underlined.

5) show that a one-base deletion (Fig. 3) and several base polymorphisms are present in the T-primer annealing site in this fragment. We suggest that these differences are responsible for the differential amplification of corresponding fragments displayed in Fig. 3. This method was also used to detect a 17-bp difference in monozygotic twins [27].

### 3.7. TDD of (CAG)<sub>n</sub> containing cDNAs

A major goal of this research has been to develop a robust targeting method that can be applied to a variety of DNAs. For instance, the same principles of targeting were used to analyze differences in (CAG)<sub>n</sub> containing subsets of expressed genes. In these experiments, mRNA isolated from different rat tissues was used as a template for TDD (Fig. 7). The used PCR primer combination included a 3 base anchoring, total. Clear differences of expressed gene fragments in different tissues can be observed.

## 4. Discussion

The semi-nested PCR targeted approach described here is the second targeting method we have developed. The first method used capture and PCR amplification for targeting, whereas the second solely uses PCR. Both methods are highly specific. However, the second method is simpler and hence potentially more suited for high throughput application, as it only requires PCR.

During the development of both methods the chosen target was (CAG)<sub>n</sub> because of the importance of these sequences in human diseases. Targeting experiments has also been conducted with (CGG)<sub>n</sub> [24], (CA)<sub>n</sub> [40], LTR's [26] and Zn-finger motifs (Storm et al., unpublished data). Ongoing work is focused on developing a repertoire of target sequences that would be useful in a wide range of applications.

Protocols like TGDD and TDD are notoriously difficult technically [41]. Furthermore, a large number of experimental variables can influence the results. A preliminary study of some variables is reported in another paper [40]. Recently we initiated experiments using genomic *Saccharomyces cerevisiae* DNA as template. This approach promises to be quite helpful in understanding variability in the protocol (Bouchard et al., unpublished results). Since the yeast genome sequence is completely known, the experimental results can be predicted and correlated with specific sequences on the genomic template.

Until now our major focus has been on genomic rather than cDNA differential display. There are a number of important applications for genomic DNA comparisons. For instance, current experiments are continuing to focus on identifying epigenetic/somatic changes

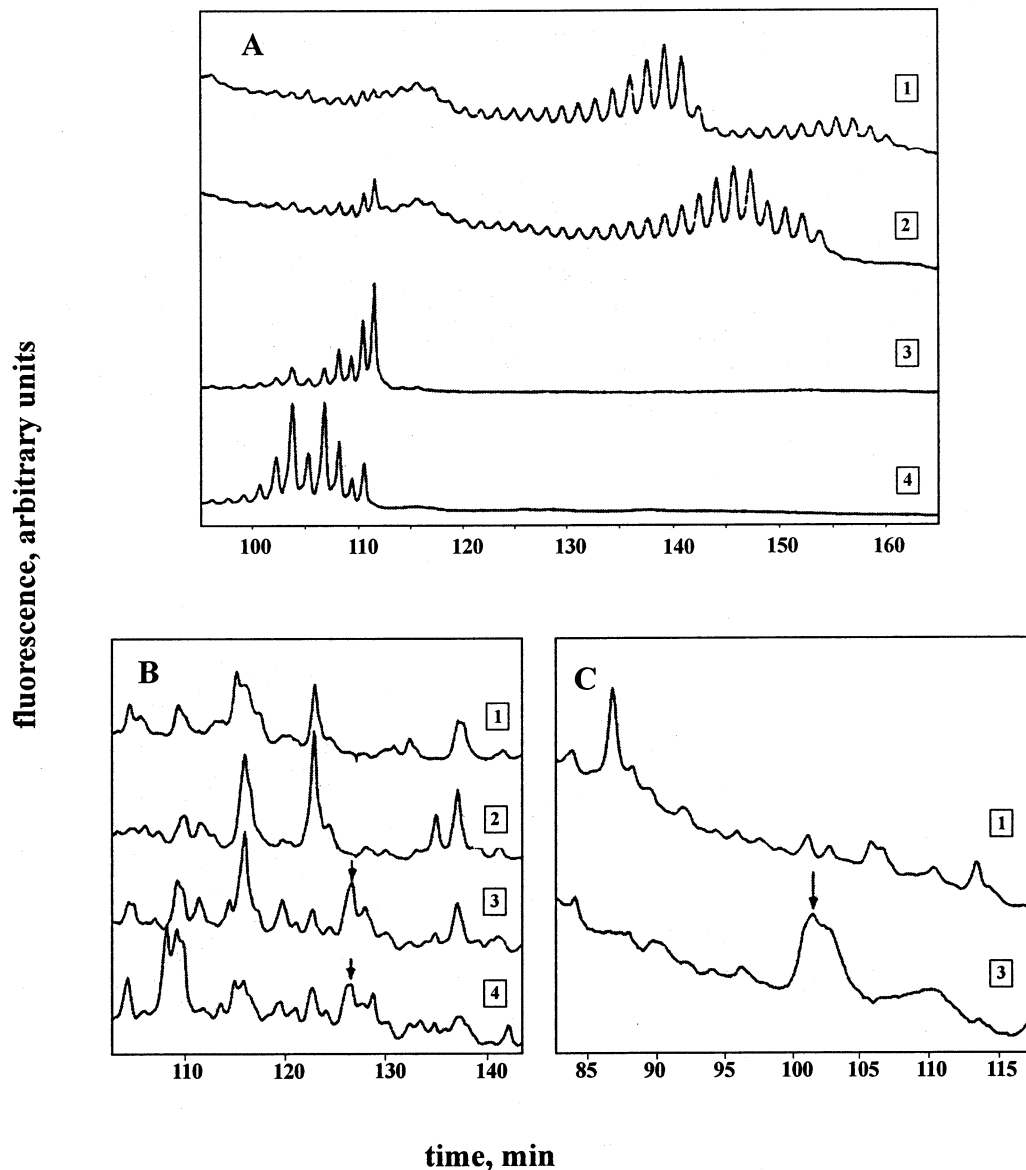


Fig. 6. TGDD of HD-alleles from HD-affected and healthy individuals (1, HD-B; 2, HD-C; 3, HD-A; 4, control). (A) HD-allele length was determined by HD specific PCR amplification using oligonucleotides 17 and 18 (Table 1). (B) TGDD of *Sau96 I* fragments using fluorescein-labeled primer (oligonucleotide 12, Table 1) in combination with oligonucleotides 3 and 16 sequentially in a semi-nested PCR reaction. The arrow indicates the expected location of the normal length HD-alleles as determined in (A). (C) The genome subset selected and shown in (B) was used as template in an HD-specific PCR reaction using a HD-primer (oligonucleotide 17) and fluorescein labeled oligonucleotide 12 (Table 1). The arrow indicates the location of a  $\sim 100$  bp fragment present in the samples with normal HD-alleles and missing in the samples with an expanded HD-allele. DNA samples were obtained from HD-affected (number 1 and 2) and healthy (number 3 and 4) individuals.

that occur during development and/or aging in different tissues and in monozygotic twins. In many instances it would be useful to obtain quantitative information about the level of differences between samples. The *S. cerevisiae* experiments mentioned above should allow us to develop quantitative methods of analysis. Ongoing experiments are applying signal processing methods to develop automated high throughput analysis of size fractionated products of the targeting procedure.

Here, the utility of a PCR based targeting method for creating libraries containing the target sequence or for comparing nucleic acid fingerprints of different samples was demonstrated. A DNA subset containing a target sequence can be used in other ways. For instance, genome subset may serve as input for subtractive hybridization methodology [21,22] or analyzed by arrays of capture probes [42]. In the latter case, analysis of array elements by mass spectrometric methods would

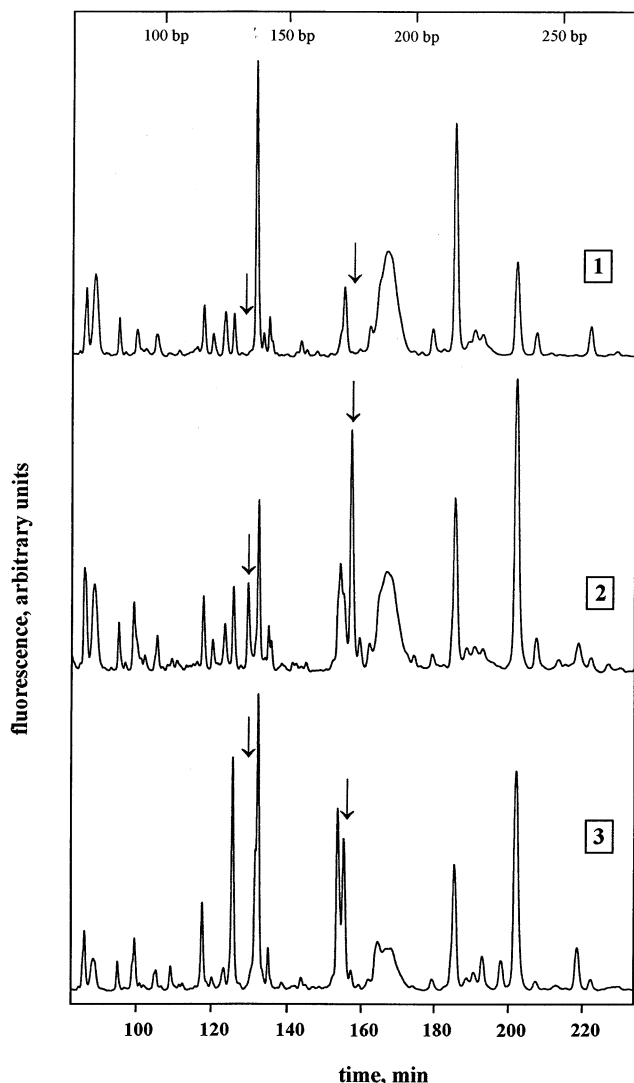


Fig. 7. Targeted differential display of double stranded cDNA isolated from heart (lane 1), medulla/pons (lane 2), and cerebellum (lane 3) of a hypertensive rat (136 mmHg MAP). cDNA was digested with restriction enzyme *RsaI* and ligated to adapter oligonucleotides. CAG-repeat containing fragments were amplified by PCR first with Cy5 labeled T-primer and an A-primer (oligonucleotides 13 and 3, respectively, Table 1) and then with the same T-primer and an anchored A-primer (oligonucleotide 6, Table 1). Fluorescent intensity versus fragment size was measured on an ALF-express DNA sequencer. The data are autoscaled in each lane by instrument software. The displayed patterns show clear differences in expressed fragments in the length range of 140 bp and at about 175 bp. Arrows ( $\downarrow$ ) indicate the polymorphic peaks.

permit multiple species to have the same capture sequence [43,44].

### Acknowledgements

We are especially grateful to Louis Keith, Arthur Pardee and David Shepro for their valuable discussions and support of this work. We thank E. Fuller

Torrey, Harry Gavras and Jim Guzella for samples. This work was supported by grants from the DOA (AIBS2154), DOE (DAM17-94-J4414) and NATO (HTECH.LG940570) to CLS and the Russian National Program to EDS.

### References

- [1] Hill AVS, Jeffreys A. Use of minisatellite DNA probes for determination of twin zygosity at birth. *Lancet* 1985;2:1394–5.
- [2] Jeffreys A, Wilson V, Thein SL. Individual-specific 'fingerprints' of human DNA. *Nature* 1985;316:76–9.
- [3] Welsh J, McClelland M. Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Res* 1990;18:7213–8.
- [4] Williams JG, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV. DNA polymorphism amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res* 1990;18:6531–5.
- [5] Caetano-Anolles G, Bassam BJ, Gresshoff PM. DNA amplification fingerprinting using very short arbitrary oligonucleotide primers. *Biotechnology* 1991;6:553–7.
- [6] Prabhu RR, Gresshoff PM. Inheritance of polymorphic markers generated by DNA amplification fingerprinting and their use as genetic markers in soybean. *Plant Mol Biol* 1994;26:105–16.
- [7] Vos P, Hogers R, Bleeker M, Reijmans M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M. AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res* 1995;23:4407–14.
- [8] Nelson DL. Interspersed repetitive sequence polymerase chain reaction (IRS PCR) for generation of human DNA fragments from complex sources. *Methods: Companion Meth Enzymol* 1991;2:60–74.
- [9] Nelson DL, Ballabio A, Victoria MF, Pieretti M, Bies RD, Gibbs RA, Maley JA, Chinault AC, Webster TD, Caskey CT. Alu-primed polymerase chain reaction for regional assignment of 110 yeast artificial chromosome clones from the human X chromosome: identification of clones associated with a disease locus. *Proc Natl Acad Sci USA* 1991;88:6157–61.
- [10] Wooster R, Cleton-Jansen A-M, Collins N, Mangion J, Cornelis RS, Cooper CS, Gusterson BA, Ponder BAJ, von Deimling A, Wiestler OD, Corneliss CJ, Devilee P, Stratton MR. Instability of short tandem repeats (microsatellites) in human cancers. *Nature Genet* 1994;6:152–6.
- [11] Huddart RA, Wooster R, Horwich A, Cooper CS. Microsatellite instability in human testicular germ cell tumors. *Br J Cancer* 1995;72:642–5.
- [12] Edwards A, Civitello A, Hammond HA, Caskey CT. DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *Am J Hum Genet* 1991;49:746–56.
- [13] Ostrander EA, Jong PM, Rine J, Duyk G. Construction of small-insert genomic DNA libraries highly enriched in microsatellite repeat sequences. *Proc Natl Acad Sci USA* 1992;89:3419–23.
- [14] Ito T, Kito K, Adati N, Mitsui Y, Hagiwara H, Sakaki Y. Fluorescent differential display: arbitrarily primed RT-PCR fingerprinting on an automated sequencer. *FEBS Lett* 1994;351:231–6.
- [15] Kandpal RP, Kandpal G, Weissman SM. Construction of libraries enriched for sequence repeats and jumping clones and hybridization selection for region-specific markers. *Proc Natl Acad Sci USA* 1994;91:88–92.
- [16] Brown J, Hardwick LJ, Wright AF. A simple method for rapid isolation of microsatellites from yeast artificial chromosomes. *Mol Cell Probes* 1995;9:53–8.

- [17] Feener CA, Boyce FM, Kunkel LM. Rapid detection of CA polymorphism in cloned DNA: application to the 5' region of the dystrophin gene. *Am J Hum Genet* 1991;48:621–7.
- [18] Ito T, Smith CL, Cantor CR. Sequence-specific DNA purification by triplex affinity capture. *Proc Natl Acad Sci USA* 1992;89:495–8.
- [19] Lench NJ, Norris A, Bailey A, Booth A, Markham AF. Vectorized PCR isolation of microsatellite repeat sequences using anchored dinucleotide repeat primers. *Nucleic Acids Res* 1996;24:2190–1.
- [20] Travis GH, Milner RJ, Sutcliffe JG. Preparation and use of subtractive cDNA hybridization probes for cDNA cloning. In: Boulton AA, Baker GB, Campagnoni AT, editors, *Neuromethods*, vol. 16, Molecular neurobiological techniques Humana, Clifton, NJ, 1989;49–78.
- [21] Lisitsyn N, Lisitsyn N, Wigler M. Cloning the differences between two complex genomes. *Science* 1993;259:946–51.
- [22] Diatchenko L, Lau YF, Campbell AP, Chenchik A, Moqadam F, Huang B, Lukyanov S, Lukyanov K, Gurskaya N, Sverdlov ED, Siebert PD. Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proc Natl Acad Sci USA* 1996;93:6025–30.
- [23] Liang P, Pardee AB. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science* 1992;257:967–71.
- [24] Broude NE, Chandra A, Smith CL. Differential display of genomic subsets containing specific interspersed repeats. *Proc Natl Acad Sci USA* 1997;94:4548–53.
- [25] Oliveira R, Broude NE, Macedo AM, Cantor CR, Smith CL, Pena SDJ. Probing the genetic population structure of *Trypanosoma cruzi* with polymorphic microsatellites. *Proc Natl Acad Sci USA* 1998;95:3776–80.
- [26] Lavrentieva I, Broude N, Lebedev Y, Lukyanov SA, Gottesman I, Smith CL, Sverdlov E. High polymorphism level of genomic sequences flanking insertion sites of human endogenous retroviral LTRs. *FEBS Lett*. 1999; in press.
- [27] Nguyen Z, Broude NE, Smith CL. Targeted genomic differential display reveals differences in genomes of monozygotic twins. 1999 (in preparation).
- [28] Siebert PDA, Chenchik A, Kellogg DE, Lukyanov KA, Lukyanov SA. An improved PCR method for walking in uncloned genomic DNA. *Nucleic Acids Res* 1995;23:1087–8.
- [29] Torrey EF, Bowler AE, Taylor EH, Gottesman II. Schizophrenia and Manic-Depressive Disorder. New York: Harper Collins, 1994:1–2 Appendix.
- [30] Wexler NS, Young AB, Tanzi RE, Travers H, Starosta-Rubin S, Penney JB, Snodgrass SR, Shoulson I, Gomez F, Ramos Arroyo MA, Penchazadeh GD, Moreno H, Gibbons K, Faryniarz A, Hobbs W, Anderson MA, Bonilla E, Conneally PM, Gusella JF. Homozygotes for Huntington's disease. *Nature* 1987;326:194–7.
- [31] Armour JAL, Neuman R, Gobert S, Jeffreys AJ. Isolation of human simple repeat loci by hybridization selection. *Hum Mol Genet* 1994;3:599–605.
- [32] Chiba H, Muramatsu M, Nomoto A, Kato H. Two human homologues of *Saccharomyces cerevisiae* SWI2/SNF2 and *Drosophila brahma* are transcriptional coactivators cooperating with the estrogen receptor and the retinoic acid receptor. *Nucleic Acids Res* 1994;22:1815–20.
- [33] Donohue PJ, Hsu DKW, Winkles JA. Differential display using random hexamer-primed cDNA, motif primers and agarose gel electrophoresis. *Meth Mol Biol* 1997;85:25–35.
- [34] Stallings RL. Distribution of trinucleotide microsatellites in different categories of mammalian genomic sequence: implications for human genetic diseases. *Genomics* 1994;21:116–21.
- [35] Gastier JM, Brody T, Pulido JC, Businga T, Sunden S, Hu X, Maitra S, Buetow KH, Murray JC, Sheffield VC, Boguski M, Duyk GM, Hudson TG. Development of a screening set for new (CAG/CTG)<sub>n</sub> dynamic mutations. *Genomics* 1996;32:75–85.
- [36] Neri C, Albanese V, Lebre A-S, Holbert S, Saada C, Bougueleret L, Meier-Ewert S, Le Gall I, Millasseau P, Bui H, Giudicelli C, Massart C, Guillou S, Gervy P, Poullier E, Rigaut P, Weissenbach J, Lennon G, Chumakov I, Dausset J, Lehrach H, Cohen D, Cann HM. Survey of CAG/CTG repeats in human cDNAs representing new genes: candidates for inherited neurological disorders. *Hum Mol Genet* 1996;5:1001–9.
- [37] Wells RD. Molecular basis of genetic instability of triplet repeats. *J Biol Chem* 1996;271:2875–8.
- [38] Demers DB, Curry ET, Egholm M, Sozer AC. Enhanced PCR amplification of VNTR locus D1S80 using peptide nucleic acid (PNA). *Nucleic Acids Res* 1995;23:3050–5.
- [39] Haddad LA, Mingroni-Netto RC, Vianna-Morgante AM, Pena SDJ. A PCR-based test suitable for screening for fragile X syndrome among mentally retarded males. *Hum Genet* 1996;97:808–12.
- [40] Storm N, Malpel S, Broude NE, Smith CL. Targeted genomic and cDNA differential display II: Factors influencing the robustness of the method. 1999 (in preparation).
- [41] Debouck C. Differential display or dismay? *Curr Opin Biotechnol* 1995;6:597–9.
- [42] O'Donnell-Maloney MJ, Smith CL, Cantor CR. The development and microfabricated arrays for DNA sequencing and analysis. *TIBTECH* 1996;14:401–7.
- [43] Fu DJ, Broude NE, Köster H, Smith CL, Cantor CR. Efficient preparation of short DNA sequence ladders potentially suitable for MALDI-TOF DNA sequencing. *Genet Anal* 1996;12:137–42.
- [44] Braun A, Little DP, Reuter D, Müller-Mysok B, Köster H. Improved analysis of microsatellites using mass spectrometry. *Genomics* 1997;46:18–23.