

Differential subtraction chain, a method for identifying differences in genomic DNA and mRNA

Jian-Hua Luo*, Janusz A. Puc, Eric D. Slosberg[1], Yao Yao[1], Jeffrey N. Bruce[2], Thomas C. Wright Jr, Michael J. Becich[3] and Ramon Parsons

Department of Pathology, [1]Herbert Irving Comprehensive Cancer Center and [2]Department of Neurosurgery, College of Physicians and Surgeons, Columbia University, 630 West 168th Street, New York, NY 10032, USA and [3]Department of Pathology, University of Pittsburgh, Pittsburgh, PA 15261, USA

Received January 27, 1999; Revised April 30, 1999; Accepted June 15, 1999

ABSTRACT

Identifying the genetic differences between two organisms or cell types has been a major goal in modern biomedical research. Recently, we developed a novel methodology that can rapidly identify the differences between two populations of DNA. This method, termed 'differential subtraction chain' (DSC), is based on a novel 'negative amplification' strategy that converts (amplifiable) tester sequences to counterpart (unamplifiable) driver sequences. The result is a double exponential elimination of amplifiable sequences in the testers, while preserving the sequences in the testers that have no counterpart in the drivers. We applied this methodology to the genome of a glioblastoma cell line. A homozygous deletion was rapidly identified. We extended this technique to identifying the unique sequences in mRNA. Two CDC25 transgene fragments were quickly identified in a *cdc25B* transgenic mouse. We also applied this methodology to systems with profound differences in mRNA expression. In a 'prostate epithelia subtracting blood cells' DSC reaction, a sample of unique gene fragments which are absent in the prostate but present in the blood were identified. Lastly, we detected rare (1 virus/100 cells) Herpes simplex virus type 2 (HSV-2) sequences in a tissue culture, indicating good sensitivity of this methodology. Overall, DSC represents a fast, efficient and sensitive method for identifying differences in genomic DNA and mRNA and can be easily applied in a variety of biological systems.

INTRODUCTION

Cloning the DNA sequence that corresponds to a genomic defect has been essential to our understanding of the genetic cause of a disease. Identifying the mRNA species that are turned on or off in specific tissues or after specific events is very useful in studies that involve embryo development, cell growth, differentiation and aging. If a unique sequence is present or a common sequence is missed in a tumor tissue when it is compared with its normal counterpart, it can be used as a tumor marker. Finding unique DNA fragments in infectious tissue helps to identify novel infectious agents (1). Numerous methods designed to identify differences in sequences have been reported. Many of them involve physical separation between testers and drivers (2–10). In the past 5 years, a PCR-based technique, called representational difference analysis (RDA), employing a differential enrichment approach to identify and enrich the differences between tested DNA samples has been developed (11). However, this method is very complicated and labor intensive. mRNA differential display (12) and RNA fingerprinting (13) by randomly primed PCR represent potentially faster and easier techniques to identify differential expression genes. However, high background and false positive results are frequently associated with these methods. These techniques also tend to be biased toward abundantly expressed sequences. Recently, a new PCR-based cDNA subtraction technique, termed suppression subtractive hybridization (SSH), was described (14). This technique used suppression PCR to preferentially amplify differential tester sequences to generate a cDNA probe library. Although this technique can dramatically enrich some differential DNA fragments, only one cycle of hybridization is permitted. Understandably, a significant background may be present.

In this report, we describe a novel PCR-based sequential subtractive hybridization method that rapidly isolates unique DNA sequences present between two tissue or cell types. This method, referred to as differential subtraction chain (DSC), employs a 'negative amplification' strategy to identify and to enrich the differences between two populations of DNA. This strategy produces fast and efficient isolation of unique tester sequences with minimal background. We have applied this method to identify genomic and mRNA differences in a variety of model systems. At the genomic level, DSC was applied to a human glioblastoma cell line and a genetic alteration in the tumor genome was rapidly identified. To detect the differences in mRNA expression, we applied DSC to a *cdc25B* transgenic mouse and rapidly identified the transgene cDNA fragments. In addition, we applied DSC to detect the differences in mRNA expression between a human prostate cell line and matched

*To whom correspondence should be addressed at present address: 3550 Terrace Street, A-721 Scaife Hall, Department of Pathology, University of Pittsburgh Medical Center, Pittsburgh, PA 15261, USA. Tel: +1 412 383 7592; Fax: +1 412 647 6676; Email: luojh+@pitt.edu

blood cells. Several novel along with two known unique gene fragments were identified.

MATERIALS AND METHODS

Materials

Breast tissue of *MMTV-cdc25B* transgenic mice were kindly provided by I. Bernard Weinstein (Columbia University). Glioblastoma cell line GL15 and matched blood cells as well as breast cancer cell lines were obtained through the Columbia University Tumor Bank. Normal primary prostate epithelial cultures were produced from organ donors as described by Krill *et al.* (15) and matched blood cells were obtained through the Western Pennsylvania Tissue Bank. Herpes simplex virus-infected cell cultures and Hep-2 cells were kindly provided by Laure Aurelian (University of Maryland). Oligonucleotides for primers/adaptors were HPLC purified and were purchased from Operon Tech (CA).

Genomic DSC

Genomic DNA of gliobastoma cell line GL15 and matched blood cells was purified by a previously described procedure (16). One to five micrograms of genomic DNA were digested with *Bgl*II (glioblastoma cell line GL15) or *Hin*dIII (HSV-2-infected Hep-2 cells) for 3–6 h. The digestion products were purified with a Qiaquick purification kit (Qiagen, CA), which was performed according to the manufacturer's recommendations. The procedure for amplicon generation was as described previously (11). Briefly, the purified restriction products were mixed with adaptor/primer (*Hin*dIII set I, AGCACTCTCCAGCCTCTCACCGCA/AGCTTGCGGTGA; *Hin*dIII set II, ACCGACGTCGACTATCCATGAACA/AGCTTGTTCATG; *Bgl*II set I, AGCACTCTCCAGCCTCTCACCGCA/GATCTGCGGTGA; *Bgl*II set II, ACCGACGTCGACTATCCATGAACA/GATCTGTTCATG) in a total volume of 20 µl TE buffer. The mixture was heated to 72°C for 3 min and cooled to 4°C in 10 min (in a PCR machine). The annealed products were ligated with Ready-to-Go ligation mixture (Pharmacia, NE) at 16°C for 30 min, which was performed according to the manufacturer's recommendations. The ligation products were purified with a Qiaquick PCR purification kit. The ligation products were mixed with all PCR reaction ingredients except *Taq* polymerase and heated to 72°C for 3 min to release the unligated oligonucleotides. *Taq* polymerase was subsequently added for 3–5 min at 72°C to fill in the sticky ends of the adaptors, followed by PCR to generate amplicons (94.5°C for 1 min, then 35 cycles of 94.5°C for 30 s and 68°C for 3 min). For hybridization, unless otherwise indicated, 10 µg of restriction enzyme-digested (6–10 h) driver DNA were mixed with 100 ng of tester DNA in 32 µl of 3× EE buffer. The mixture was heated to 98–100°C for 5 min and 8 µl of 5 M NaCl was added. The mixture was then incubated at 67°C, for 14–20 h when genomic amplicons were used, for 2–3 h when the λ amplicon was used. The reannealed products were purified with a Qiaquick PCR purification kit. The purified DNA was resuspended in 50 µl of 1× mung bean nuclease buffer (New England Biolabs, MA) with 10 U of mung bean nuclease at 30°C for 25–30 min. The mung bean nuclease was subsequently removed with a Qiaquick PCR purification kit. An aliquot (5–10%) of the purified DNA products was taken for PCR to examine subtraction efficiency. The remainder was reheated to 98°C for second round hybridization. No additional driver was added after the first round of DSC. The PCRs were performed using gold *Taq* polymerase (Perkin Elmer) under the following conditions: 95°C for 1 min, 95°C for 15 s, 68°C for 3 min for 35–43 cycles. Southern blot hybridizations were performed according to standard procedures (17). The primer sequences for *p16* were GGAAATTGGAAACTGGAAGC/TCTGAGCTTTGGAA GCTC. The primer sequences for *L1* were GATAGGTTGCATGGCCTACC/TCAAGTGG TTCC-TGTGCTGG.

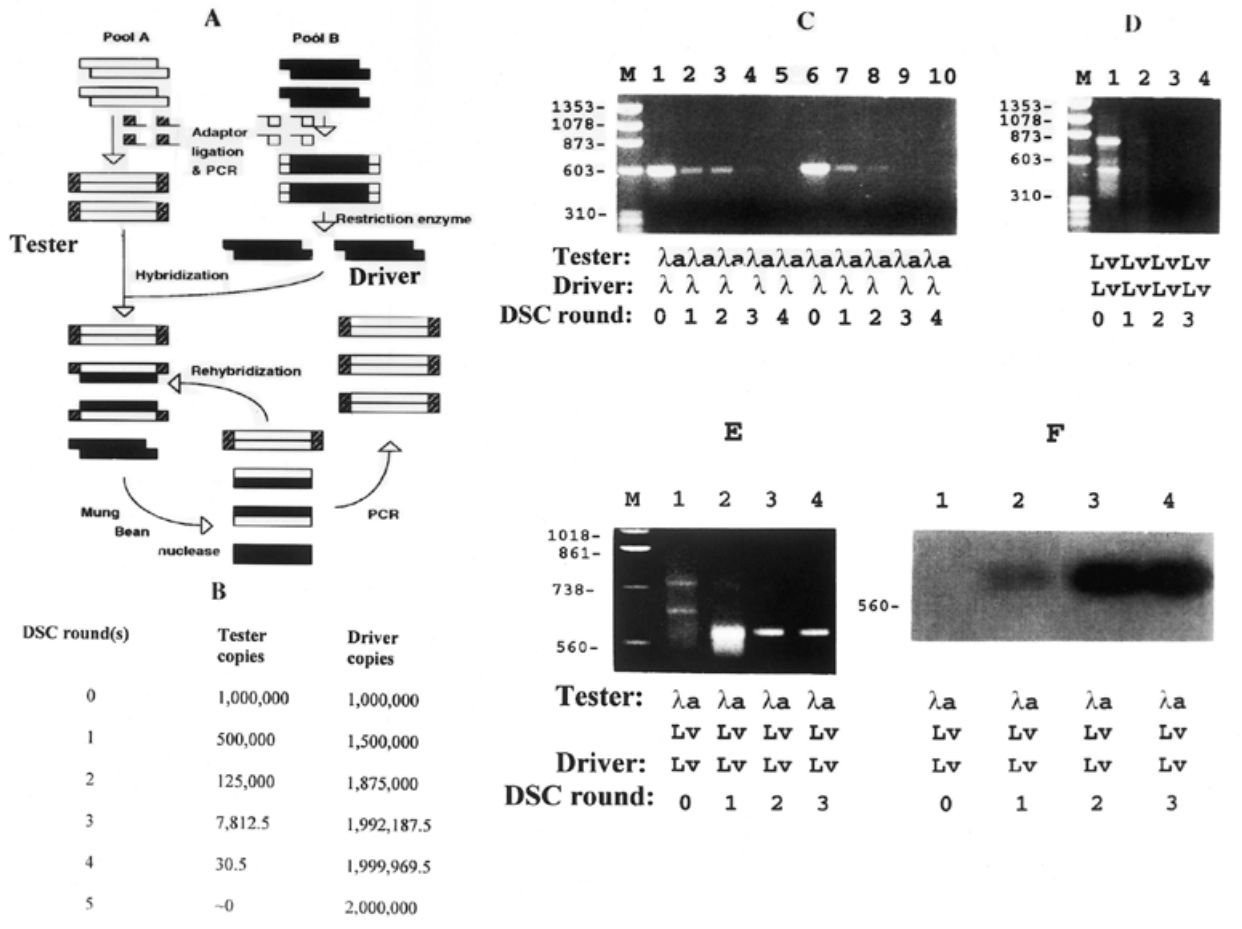
cDNA DSC

mRNA from tumor cells was purified using oligo(dT)–agarose columns (Pharmacia). The purification procedure was performed according to the manufacturer's recommendations. Double-strand cDNA synthesis was performed using the template switch technique (Smart™ cDNA library construction kit; Clontech, CA). The PCR-amplified cDNAs were then digested with *Dpn*II and ligated similarly as described above with two separate sets of primer/adaptor (set I: BamIa, ATGAAGTGCACCCTACGATTCGAG, BamIb, pGATCCTCGAATCGTAGTGGTGCACT; set II: BamIIa, ATGAGACATGTTTCGTAGCCTAGG, BamIIb, pGATCCCTAGGCTACGAAACATGTC). The ligation products were purified with a Qiaquick PCR purification kit and subject to PCR amplification using *Taq* polymerase (Perkin Elmer) under the following conditions: 72°C for 3 min, 94°C for 1 min, 68°C for 3 min for 30–35 cycles. For cDNA DSC, hybridization time was reduced to 10–14 h in each round. Northern blot hybridizations were performed according to a previously described procedure (18). The primer sets for the JBL series were as follows: JBL1a, ATCGGCACCAGCACGATGAGGTTG/JBL1b, GCGCTATCAGCATTCTTCGATGAG; JBL2a, TCTCTACCTGAGGGCAAGAGTGTG/JBL2b, CTCATAGTCTGG GTCGGGAACAGG; JBL3a, TTGCCGATGCTGTCGAGCAGGTTG/JBL3b, GGGCGTTGTGTAACACCAGAAAGG; JBL4a, GACTTCAGCTCATTGTCCACATAG/JBL4b, AGCATCCCTTGCAGTGCACTATGG; JBL5a, GTCACTGAGGTCAAGACCAAGACC/JBL5b, TCTCGACCAGCTTGACATCACAGG; JBL6a, AGCCCAAGCTCCACAAGGCTTTCC/JBL6b, TTCCACGCGAGCTAGGCCCTGTTG; JBL7a, GCTTGTGACTGCCAAGCTTGAC/JBL7b, AGGCCAGCGGGACCTGTATTCTGG; JBL8a, GCTCGCAAGCTTGCAGAACGACTG/JBL8b, GTCCGCGTTGAAAGTGAGCTATCG; JBL9a, TTGCGCGTAACGCATCGGATTATG/JBL9b, AAACGCGGGA GCGACCCGATGAGA; JBL10a, CATGTCTCGCACATCTGGAGGGAC/JBL10, GAGGAATCCTTGAATCACTGCGCC. Primers specific for ubiquitin were as follows: GACGCAAACATGCAGATCTTTGTG/AATGAAAGGGACACTTTATTGAGG.

RESULTS

Principles of DSC

Identification of genomic deletions or novel infectious agents has usually been accomplished by subtractive hybridization. However, due to the high complexity of eukaryotic genomes and the low enrichment of target sequences, this method was found inefficient for subtracting genomic DNAs from higher eukaryotes. This limits the application of this method to the identification of relatively long and relatively abundant



B

DSC round(s)	Tester copies	Driver copies
0	1,000,000	1,000,000
1	500,000	1,500,000
2	125,000	1,875,000
3	7,812.5	1,992,187.5
4	30.5	1,999,969.5
5	~0	2,000,000

Figure 1. Differential subtraction chain. (**A**) Schema of differential subtraction chain. Pool A DNA is used as testers and Pool B as drivers. Pool A and B represent restriction enzyme-digested DNA fragments. These fragments are ligated with separate sets of adaptors/primers. Subsequently, the ligation products are amplified with primers specific to adaptors/primers to generate amplicons (adaptor-tagged restriction DNA fragments). For pool B DNA, amplicons are digested with restriction enzyme to remove the adaptors to become drivers before hybridization. In the DSC reaction, mixtures of testers and drivers are denatured and hybridized. Subsequently, the hybridization products are treated with mung bean nuclease to remove the single-stranded primers. After mung bean nuclease is removed, DNA is subjected to denaturation and rehybridization. An aliquot (5–10%) of DSC reaction stock is taken for PCR after each round of DSC. These PCRs are not required for the DSC reaction and are for quality control purposes. (**B**) Mathematical model of 'negative amplification'. Hypothetical results of DSC, assuming the starting tester/driver ratio is 1. The survival number(s) of tester after each round of DSC follows the probability of tester to self-anneal in the preceding round, which is 1/2 before round 1 of DSC, 1/4 after round 1, 1/16 after round 2, 1/256 after round 3, 1/65536 after round 4,…, and $1/2^{2^n}$ after *n* rounds. Therefore, the survival copy(s) of tester can be predicted by the equation $Y_e = y_{A^{2^n}}$, where Y_e is the probable copy number(s) of survived tester after *n* round(s) of DSC, *y* is the total copy number(s) of a tester species and its relevant driver and *A* is the probability of the tester(s) self-annealing before the first round of DSC. Similarly, the number(s) of tester converted to driver in each round of DSC follows the probability of tester cross-annealing with driver, which is 1/2 before round 1 of DSC, 3/4 after round 1 of DSC, 15/16 after round 2, 255/256 after round 3, 65535/65536 after round 4,…, and $1 - 1/2^{2^n}$ after *n* rounds. The copies of driver that survive each round of DSC can be predicted by the equation $Y_{or} = y(1 - A^{2^n})$, where Y_{or} denotes total copy number(s) of driver after *n* round(s) of DSC. The probability for complete elimination of a tester species can be assessed $(1 - y_{A^{2^n}})$. (**C**) Subtraction efficiency of DSC on the λ amplicon. Five nanograms of 610 bp λ amplicon (λa) were subtracted by 3.9 (lanes 1–5) or 39 μg (lanes 6–10) of total λ DNA (*Hin*dIII restricted). The DSC products were amplified by PCR after 0 (lanes 1 and 6), 1 (lanes 2 and 7), 2 (lanes 3 and 8), 3 (lanes 4 and 9) and 4 (lanes 5 and 10) rounds of DSC. (**D**) Subtraction efficiency of DSC in human genomic amplicons. One hundred nanograms of human liver genomic amplicon testers were subtracted in a DSC reaction in the presence of 10 μg of liver amplicon drivers from the same tissue. The DSC products were amplified by PCR after 0 (lane 1), 1 (lane 2), 2 (lane 3) and 3 (lane 4) rounds of DSC. (**E**) Identifying the presence of λ DNA sequence between two DNA pools using DSC. Fifty picograms of λ DNA were mixed with 1 μg of human liver genomic DNA to generate mixed amplicon testers. One hundred nanograms of these testers were subtracted by 10 μg of liver genomic amplicon drivers. The DSC products were amplified by PCR after 0 (lane 1), 1 (lane 2), 2 (lane 3) and 3 (lane 4) rounds of DSC. (**F**) Southern blot hybridization of (E) with ^{32}P-labeled λ DNA probe.

sequences. To reduce the complexity of genomic sequences in hybridization, we adopted the concept of genomic representational sampling proposed by Lisitsyn *et al.*, in which short fragments of endonuclease-digested genomic DNAs are ligated with adaptors and amplified to produce 'amplicons' (11; Fig. 1A). Presumably, the genomic representation by amplicons would reduce the complexity of sequences by 80- to 100-fold and dramatically reduce the time needed to perform hybridization (11). However, once the amplicons were generated, we abandoned the differential enrichment approach. Instead, we developed a 'negative amplification' strategy to eliminate homologous sequences present in two different DNA pools. The principle of our strategy is based on the assumption that the maximum efficiency of a subtractive process depends on

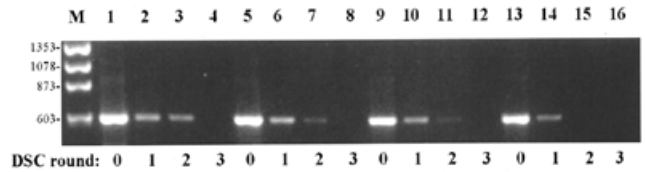


Figure 2. Synchronization of subtraction of DSC. The λa amplicon at 50 (lanes 1–4) and 5 ng (lanes 5–8) and 500 (lanes 9–12) and 1 pg (lanes 13–16) was used as tester in DSC reactions and was subtracted by 100-fold of the corresponding λa driver. The λa tester was amplified after 0 (lanes 1, 5, 9 and 13), 1 (lanes 2, 6, 10 and 14), 2 (lanes 3, 7, 11 and 15) and 3 (lanes 4, 8, 12 and 16) rounds of DSC. Templates from lanes 1, 5 and 9 were diluted with distilled water to produce 1000-, 100- and 10-fold dilutions, respectively, before PCR amplification.

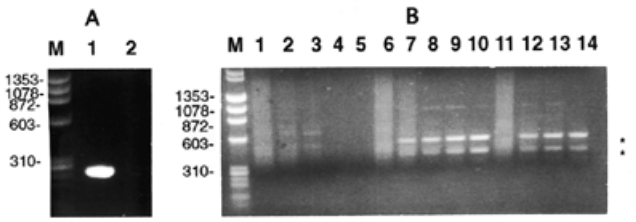


Figure 3. Detecting the differences in rare DNA sequences between two populations of DNA. (**A**) PCR products of primers specific for HSV-2, from HSV-2-infected (lane 1) and uninfected (lane 2) Hep-2 cells. (**B**) Detection of HSV-2 fragments using DSC. Hep-2 cells mock infected (lanes 1–5) or infected with HSV-2 at low (0.01 m.o.i.) (lanes 6–10) and high (5 m.o.i.) (lanes 11–14) multiplicity of infection. Cells were harvested 4 h after infection. *Hin*dIII amplicons were generated. For mock or low multiplicity of HSV-2 infection, the starting quantities of the uninfected or the HSV-2-infected Hep-2 cell testers were 7 µg and were subtracted by 33 µg of uninfected Hep-2 cell drivers. For high multiplicity of infection, the starting quantity of HSV-2-infected Hep-2 cell testers was 100 ng and was subtracted by 10 µg of uninfected Hep-2 cell drivers. The DSC products were amplified after 0 (lanes 1, 6 and 11), 1 (lanes 2, 7 and 12), 2 (lanes 3, 8 and 13), 3 (lanes 4, 9 and 14) and 4 (lanes 5 and 10) rounds of DSC. The identities of the 580 (*) and 430 bp (*) DSC products were confirmed to be HSV-2 DNA fragments through sequencing.

the contrast ratio of the DNA that is used to subtract (driver, non-adaptor-tagged amplicon) to the DNA being subtracted (tester, adaptor-tagged amplicon). In subtractive hybridization (2–10) or RDA (11), a high ratio between the drivers and testers is achieved through repeated addition of drivers and sequential dilution of testers in each cycle of subtractive hybridization. Unfortunately, this may result in the loss of rare sequences as well as an increase in time spent and in number of steps in the procedure. In DSC a high ratio between drivers and testers is obtained through a process that converts testers to homologous drivers after each subtractive round. This 'negative enrichment' process is accomplished by removing the primers from both ends of the testers with mung bean nuclease after they cross-annealed to drivers (Fig. 1A). In the subsequent round of subtractive hybridization, the testers without primers will become drivers and serve to convert more testers to drivers from the tester population. The result is a double exponential decrease in the copy number(s) of tester species that have homologous counterparts in the driver population and a corresponding increase in the copy number(s) of the driver species (see Fig. 1B for a mathematical evaluation). We term this process 'negative amplification'. Presumably, if 10^9 copies of a target sequence present in one population is to be eliminated ($Y_e < 1$; see Fig. 1B for definition) in order to achieve subtraction, then in the presence of 99-fold ($y = 10^{11}$, $A = 1/100$) copies of the corresponding driver it would take 3 ($n = 2.46$) rounds of negative amplification to achieve complete elimination of this tester. This negative amplification process achieves two goals. (i) By complete removal of the primers from the unwanted sequences, it eliminates any possible amplification while preserving the rare DNA sequences. This will improve the subtraction sensitivity and efficiency and produce cleaner products (less false positive results). (ii) By bypassing the need for positive enrichment processes between hybridizations, it dramatically simplifies and speeds up the subtractive process.

To demonstrate the subtraction efficiency of DSC, a 560 bp fragment of *Hin*dIII-digested λ DNA was ligated with an adaptor/primer and amplified to generate a tester (λa). More than 7.7×10^9 (5 ng) copies of this amplified fragment were used in a DSC subtraction in the presence of 10- or 100-fold of the corresponding DNA fragment of λ DNA (driver). As shown in Figure 1C, complete elimination of the tester was achieved after four rounds of DSC when 10-fold copies of driver were applied, while only three rounds of DSC were required to reach a similar result when 100-fold copies of driver were used. These results fit well with the mathematical model shown in Figure 1B: when 10-fold copies of driver over tester were applied, ~395 copies of tester survived after three rounds of DSC. The minimum number of copies of tester that can survive four rounds of DSC is 4.18×10^{15} (when $Y_e = y_A{}^{2^n} = 1$ and $y_0 = y_A$, where y_0 denotes tester copy number before the first round of DSC, then $y_0 = 1/A^{2^n - 1}$). When 100-fold copies of driver were applied, the minimum copies of tester that were needed to survive three rounds of DSC were 1.07×10^{14}. Clearly, complete subtraction of the λ amplicon should be achieved within four rounds of DSC if 10-fold driver is used and within three rounds if 100-fold is used. To evaluate the utility of DSC in human genome subtraction, amplicons generated from genomic DNA of a human liver (100 ng) were used as testers and were subtracted by 100-fold copies of drivers generated from the same tissue. No tester survived two rounds of DSC (Fig. 1D), reflecting lower copies of individual tester species. The ability of DSC to identify differences between two pools of DNA was evaluated by mixing minute quantities of λ DNA with human liver genomic DNA to generate mixed testers and subtracting the mixture in the presence of 100-fold copies of liver drivers. After two rounds of DSC, only the λ amplicon was amplified (Fig. 1E and F).

In DSC, with the hybridization kinetics and double exponential nature of tester to driver conversion, the more abundant a tester species, the greater the number which will cross-hybridize with drivers and be converted to drivers. Thus, elimination of high abundance species is faster than that of low abundance ones. This normalization (equalization) function of DSC can be predicted mathematically, since the conversion rate ($Y_{convert}$) of a tester to driver is expressed by a linear regression relationship with the total abundance of that specific tester–driver (y) in a given round (n). $Y_{convert} = ye(A - A^{2^n}) + yd(A - A^{2^n})$, where $Y_{convert}$ is the probable copy number(s) of tester conversion to driver after n rounds of DSC, ye is the total copy number(s) of a tester species, yd is the total copy number(s) of the relevant driver and

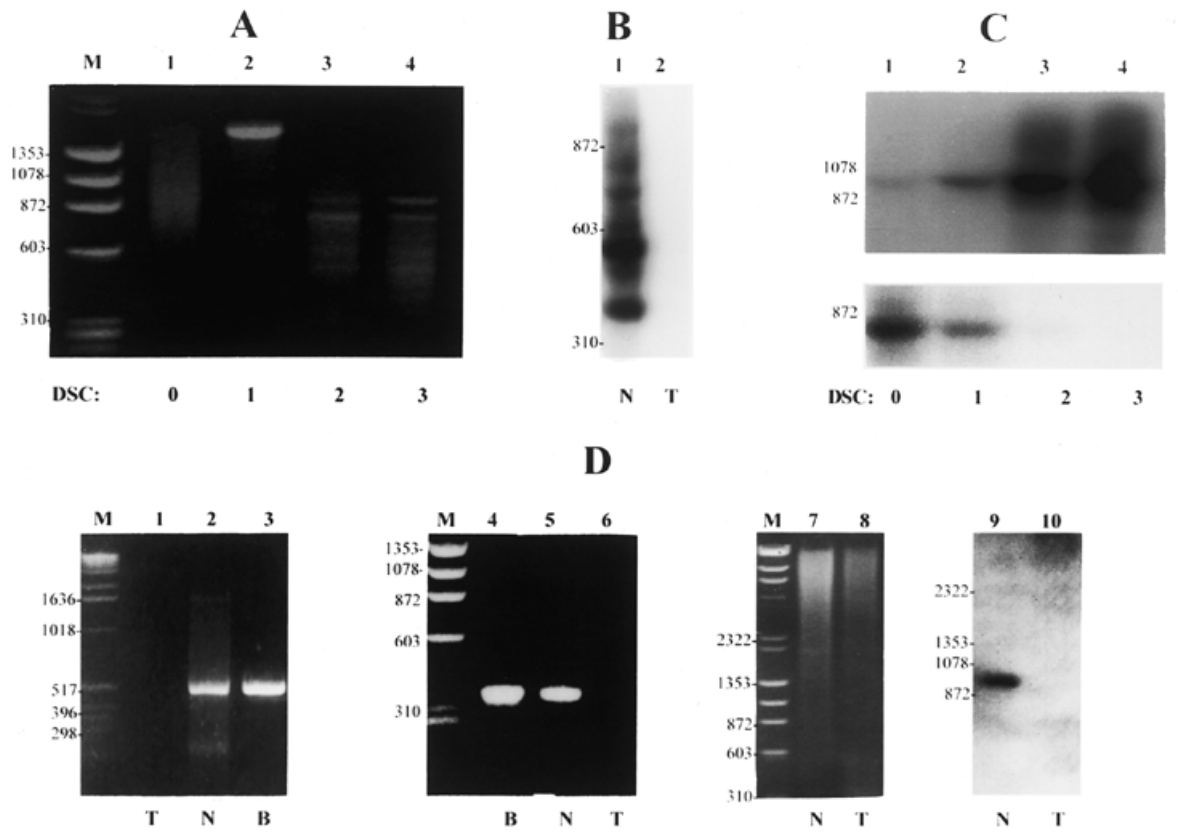


Figure 4. Identification of genomic DNA sequences missing in glioblastoma cell line GL15. (**A**) DSC products of *Bgl*II (lanes 1–4) testers from matched normal cells when 100-fold of respective gliablastoma cell drivers were applied, obtained after 0 (lane 1), 1 (lane 2), 2 (lane 3) and 3 (lane 4) rounds of DSC. (**B**) Southern blot hybridization of amplicons with DSC products. Round 3 DSC product *Bgl*II amplicons were used as probes to hybridize to respective amplicons generated from normal (N) or tumor (T) tissue. (**C**) Enrichment of *L1* sequence from normal amplicons by DSC. Lanes 1–4 represent Southern blot hybridization on DSC products of *Bgl*II testers from rounds 0 (lane 1), 1 (lane 2), 2 (lane 3) and 3 (lane 4) of DSC, with the PCR products of the *L1* (upper panel) and *C15* (lower panel) clones as probes. (**D**) Homozygous deletion of the *L1* probe in GL15. Lanes 1–3 represent PCR products of primers specific for exon 2 of the *p16* gene from GL15 cells (lane 1, T), normal tissue (lane 2, N) and BAC clone *59I23* (lane 3, B). Lanes 4–6 represent PCR products of primers specific for the *L1* probe, from BAC clone *59I23* (lane 4), normal tissue (lane 5) and GL15 cells (lane 6). Lanes 7–10 represent Southern blot hybridization on *Bgl*II-digested genomic DNA from normal (lanes 7 and 9) and GL15 (lanes 8 and 10) cells, with *L1* as probe.

A is the probability of tester self-annealing before the first round of DSC. Because of the enormous normalization ability of DSC, subtraction of a tester species ranging from 1.03×10^7 to 1.07×10^{14} in copy number will synchronize when 100-fold copies of driver are applied, i.e. a tester with a copy number within this range will survive the second round of DSC but not the third round. To demonstrate this normalization function of DSC, serially diluted λa tester was mixed with 100-fold copies of λa driver in DSC reactions. As demonstrated in Figure 2, λa tester starting with 50 ng (7.7×10^{10}), 5 ng (7.7×10^9) and 500 pg (7.7×10^8) survived two rounds of DSC reaction but not three rounds, while λa tester starting with 1 pg (1.54×10^6) did not survive the second round of DSC. These results clearly confirm our mathematical prediction and indicate that DSC subtracts effectively regardless of variations in abundance of individual tester species.

To demonstrate the sensitivity of DSC to detect differences in rare sequences, Hep-2 cell cultures were infected with low or moderate multiplicity of infection units (m.o.i.) of Herpes simplex virus type 2 (HSV-2) (0.01 or 5 m.o.i., respectively). Shortly after infection, DNA was extracted from the cultures. As demonstrated in Figure 3, after two rounds of DSC, two DNA fragments from both low and moderate multiplicity of infection were identified, which were confirmed to be HSV-2 sequences by sequencing. In addition, by mixing serially diluted artificial λa tester with hepatic cell genomic testers and subtracting with hepatic cell drivers, DSC scored positive on the λa tester starting with less than 100 copies (data not shown).

Application of DSC to detect genomic homozygous deletion

To test the ability of DSC to identify genetic defects in human diseases, a human glioblastoma cell line GL15 with a known homozygous deletion encompassing *p16*/*p15* (CDKNA/CDKNB) was used as our test model. Amplicons of GL15 and matched blood DNA were generated by *Bgl*II digestion. To identify DNA deleted in the tumor, the normal amplicons were used as the testers and the tumor amplicons were used as the drivers. As shown in Figure 4A, after three cycles of subtraction a distinctive band pattern was identified. To test whether these

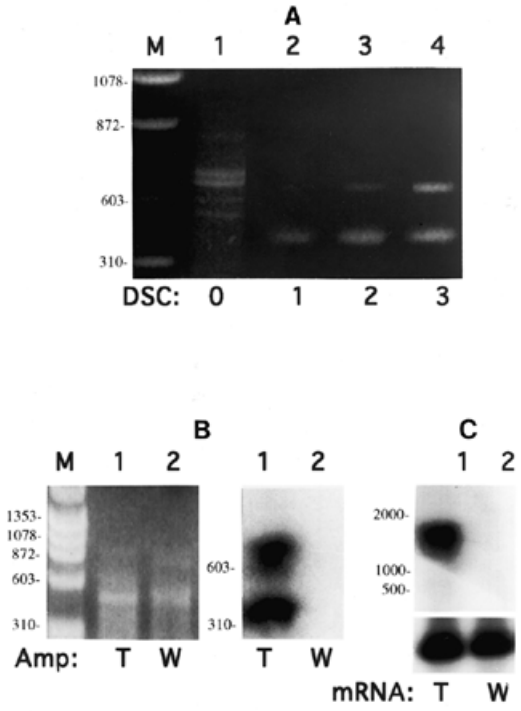


Figure 5. Detection of *cdc25B* cDNA fragments in a *MMTV-cdc25B* transgenic mouse using DSC. (**A**) Agarose gel electrophoresis of DSC products of *Dpn*II cDNA amplicon testers from the transgenic mouse when excess *Dpn*II cDNA amplicon drivers from wild-type mouse were applied, obtained after 0 (lane 1), 1 (lane 2), 2 (lane 3) and 3 (lane 4) DSC rounds. (**B**) Southern blot hybridization on *Dpn*II cDNA amplicons by DSC products. Round 3 DSC products of *Dpn*II cDNA testers from the *MMTV-cdc25B* transgenic mouse were used as probes to hybridize to cDNA amplicons generated from transgenic (lanes 1 and 3) and wild-type (lanes 2 and 4) mice. (**C**) Northern blot analysis of the DSC probe. A 650 bp DSC product was used as a probe to hybridize to total mRNA extracted from *cdc25B* transgenic (lane 1) and wild-type mice (lane 2). The lower panel represents hybridization with probes from glycerol 3-phosphate dehydrogenase. Similar results were obtained with the 350 bp DSC product.

amplified DSC products were only present in the normal genomic amplicons, round 3 DSC products were used as a probe to perform Southern blot analyses on amplicons from GL15 and its paired normal DNA. Several distinct bands that corresponded to the DSC products were identified in the normal amplicons (Fig. 4B), while no distinct hybridization bands were seen in the tumor counterpart, indicating that these DSC products were not represented in tumor amplicons. Ten clones of DSC products after round 3 were randomly selected and sequenced. Primers specific for four unique DSC product sequences were made to examine whether these DSC products represented homozygous deletion in tumor cells. One of the four DSC products examined was found to be homozygously deleted in the tumor genome (Fig. 4D). This product, designated *L1*, was present in a bacterial artificial chromosome clone (59I23) that contains *p16*, arguing that *L1* and *p16* are within the same deletion. *L1* was greatly enriched during DSC subtraction, in contrast to de-enrichment of probe *C15*, which is shared by both normal and tumor amplicons (Fig. 4C). In addition, a panel of 25 breast cancer cell lines was screened with *L1* primers (data not shown). Three of the four lines harboring homozygous deletions of *p16* were also deficient in *L1*. *L1* could be amplified in all lines containing *p16*.

The majority of DSC products are not homozygously deleted, but appear to be missing from the GL15 tumor amplicons. As has been noted by Lisitsyn and Wigler (11), many of the products of effective subtraction are due to restriction length polymorphisms that are deleted by loss of heterozygosity. Additionally, point mutations that affect *Bgl*II sites can lead to effective subtraction.

Application of DSC to detect differences in mRNA expression

With slight modification, DSC can be used to identify mRNA sequences that are uniquely present in one organism compared with another in 2–3 days. The human-derived *MMTV-cdc25B* (19) transgenic mouse was chosen as a model system to test the usefulness of DSC in identifying unique mRNA sequences. To generate cDNA amplicons, a 4 bp cutter restriction enzyme *Dpn*II that produces a 4 bp 5'-overhang was selected. On average, this enzyme will cut every 256 bp. Presumably, most of the cDNA will be cut at least twice so that amplicon fragments can be generated (20). As shown in Figure 5A, excessive *Dpn*II-restricted cDNA amplicons from wild-type mouse were used as drivers to subtract the amplicons from the *cdc25B* transgenic mouse. After three rounds of DSC subtraction, two distinct (650 and 350 bp) DNA fragments were amplified. Southern blot hybridization with these DSC products confirmed that they were only present in transgenic cDNA amplicons (Fig. 5B). These two DSC products were subsequently cloned and sequenced and confirmed to be fragments of *cdc25B* cDNA. These cloned DSC products were then used as a probe to perform northern blot hybridization on mRNA extracted from wild-type and *cdc25B* transgenic mice (Fig. 5C). The results confirmed that these probes are unique to the transgenic mouse.

To test the utility of DSC in systems with profound differences in mRNA expression, cDNA amplicons from a primary human prostatic epithelium culture (N277) were used as drivers to subtract counterparts from the blood cells of the same individual. As shown in Figure 6A, after three rounds of DSC, multiple amplified tester fragments emerged. To verify whether these

Table 1. Differentially expressed gene fragments of blood cells from prostate

Clone	Gene sequence	No. of clones
BDP1	Unknown	1
BDP2	CD3 ε chain	3
BDP3	Unknown	1
BDP4	Unknown	4
BDP5	T cell receptor α chain	2
BDP6	Unknown	1
BDP7	Unknown	2
BDP8	Unknown	1
BDP9	Unknown	3
BDP10	Unknown	2

BDP, blood differential from prostate gene fragment.

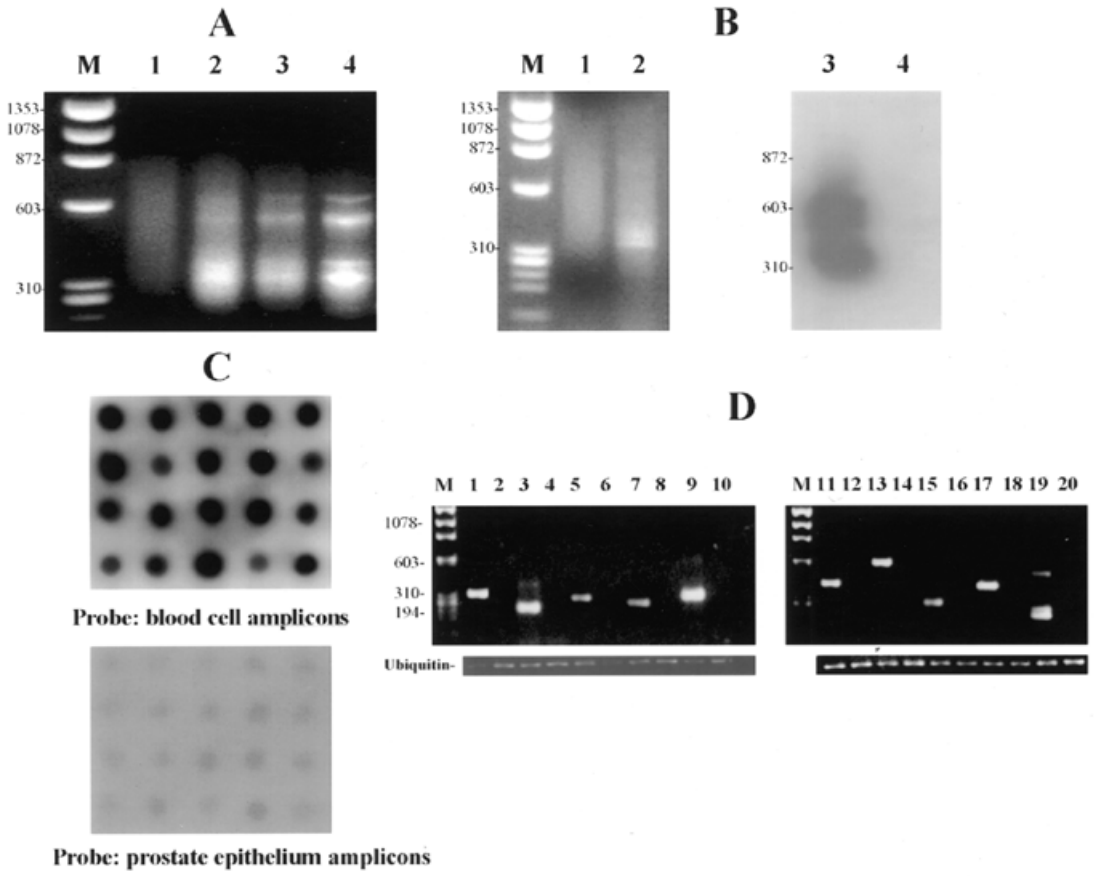


Figure 6. Detecting qualitative differences in mRNA expression between human prostatic epithelium and blood cells. (**A**) Agarose gel electrophoresis of DSC products after 0 (lane 1), 1 (lane 2), 2 (lane 3) and 3 (lane 4) rounds of subtraction, when 100 ng of blood cell amplicon testers were subtracted by 10 µg of prostate epithelium amplicon drivers. (**B**) Southern blot hybridization of amplicons from blood cells and prostate epithelium with DSC products. Round 3 DSC products from (A) were used as probes to hybridize amplicons generated from blood cells (lanes 1 and 3) and prostate (lanes 2 and 4) epithelium of the same individual. (**C**) Colony hybridization of clones of round 3 DSC products. Twenty randomly picked colonies from TOPO-TA cloning of round 3 DSC products of (A) were hybridized with amplicon probes from blood cells and prostate epithelium. (**D**) RT–PCR of mRNA from blood cells and prostate epithelium with primers specific for the BDP gene. Purified mRNA from blood cells (lanes 1, 3, 5, 7, 9, 11, 13, 15, 17 and 19) and prostate epithelial cultures (lanes 2, 4, 6, 8, 10, 12, 14, 16, 18 and 20) were reversed transcribed. PCRs were performed using primers specific for BDP1 (lanes 1 and 2), BDP2 (lanes 3 and 4), BDP3 (lanes 5 and 6), BDP4 (lanes 7 and 8), BDP5 (lanes 9 and 10), BDP6 (lanes 11 and 12), BDP7 (lanes 13 and 14), BDP8 (lanes 15 and 16), BDP9 (lanes 17 and 18) and BDP10 (lanes 19 and 20). The bottom panel represents RT–PCR of the ubiquitin gene fragment.

products represent fragments of genes present only in blood cells, the round 3 DSC products were used as probes to perform Southern blot hybridization on amplicons from blood cells and prostate epithelium. An intense hybridization reaction was seen with the blood cell amplicons, while a negative reaction was found with the prostate epithelial counterpart (Fig. 6B). The mixture of round 3 DSC products was subsequently cloned. Twenty colonies were randomly picked and sequenced. As shown in Table 1, 10 species of unique sequences were identified. Two of these were the T cell receptor ε and α chains. Surprisingly, the rest of the fragments represent previously unknown sequences. To further investigate the uniqueness of these fragments in the blood cells, RT–PCR on mRNA from prostate epithelium and blood cells was performed using primers specific to these sequences. As shown in Figure 6D, no PCR product was identified in mRNA from prostate epithelium, while PCR products with the appropriate sizes were seen in the blood cell counterpart.

DISCUSSION

The DSC method described in this report offers a good alternative to RDA and conventional subtractive hybridization as a method for identifying unique DNA sequences, because DSC represents a considerably easier (75% fewer steps) and faster procedure (2–3 days compared with weeks or months). DSC also offers excellent subtraction efficiency. This is because with DSC the deprimerized testers from the previous subtraction serve as drivers in the following cycles. DSC appears to have good sensitivity that facilitates detecting differences in rare DNA sequences between two DNA pools. This is because DSC tolerates low starting driver/tester ratios and the starting quantity of rare testers can be increased dramatically. Furthermore, the DSC subtraction series are performed without tester dilution. When uniquely expressed mRNA species are desired, DSC offers several technical advantages over differential display and RNA fingerprinting (12,13). These include: a clean background;

being non-radioactive; being free of manual gel excision; having the capability of evaluating total mRNA in one shot; while retaining similar speed. In comparison to SSH, DSC has the advantage of producing nearly pure differential probes for mRNA expression because of its capability to undergo multiple cycles of subtraction and the nature of negative amplification. This will save time in screening cDNA probes. The wide variance in abundance of individual mRNA species does not appear to affect the effective subtraction of DSC. This can be attributed to two reasons. (i) Similarly to RDA and SSH, DSC employs two PCR procedures to generate mRNA amplicons (one for cDNA generation, the other for amplicon production). Significant normalization of gene fragment representations in amplicons occurs (data not shown). (ii) Because of the enormous synchronization function of DSC, it normalizes the wide range of abundances of testers quickly. In addition, the high subtractive power of DSC due to the double exponential nature of 'negative amplification' also makes it easy to overcome the problem of variation of species abundance. Theoretically, if a high driver/tester ratio (e.g. 100:1) is employed, DSC is capable of subtracting away more than 69 μg of a 610 bp tester within three cycles. In this regard, DSC is best suited to detecting qualitative differences in mRNA expression. Thus, techniques designed to address the relative abundance of mRNA expression, such as serial analysis of gene expression (SAGE) (21), will be a good complement to DSC.

Because of high speed, good sensitivity and low background, DSC will greatly facilitate studies in locating genomic defects, detecting polymorphisms and identifying specific genes involved in cell growth, differentiation and aging, as well as probing for the presence of infectious agents. Since DSC typically takes less than 3 days to generate results, it has the potential to be used as a routine method to search for genetic markers associated with cancer cases and to identify infectious agents.

ACKNOWLEDGEMENTS

We thank I. Bernard Weinstein, Michael Shelanski and Steven Brown for constructive comment on the manuscript. J.H.L. thanks Patricia Lanzano for technical help on Southern blot analysis and Drs Laure Aurelian and Cindy Smith for providing the HSV-2-infected Hep-2 cells. We thank Judith Stoner for preparing primary prostate tissue cultures, Petrina Deflavia for providing matched organ donor blood cells and Barbara Siket for editing this manuscript. The laboratories of T.W., M.J.B. and R.P. made equal contributions to this work.

REFERENCES

1. Chang,Y., Cesarman,E., Pessin,M.S., Lee,F., Culpepper,J., Knowles,D.M. and Moore,P.S. (1994) *Science*, **266**, 1865–1869.
2. Hendrick,D.M., Cohen,D.I., Nielsen,E.L. and Davis,M.M. (1984) *Nature*, **308**, 149–153.
3. Davis,M. (1984) *Proc. Natl Acad. Sci. USA*, **81**, 2129–2194.
4. Duguin,J.L. and Minauer,M.C. (1990) *Nucleic Acids Res.*, **18**, 2789–2792.
5. Hara,E., Kato,T., Nakada,S., Sekiya,L. and Oda,K. (1991) *Nucleic Acids Res.*, **19**, 7097–7104.
6. Sargent,T.L. and Dawid,I.B. (1983) *Science*, **222**, 135–139.
7. Lamar,E.E. and Palmer,E. (1984) *Cell*, **37**, 171–177.
8. Kunkel,L.M., Monaco,A.P., Middlesworth,W., Oches,H.D. and Latt. S.A. (1985) *Proc. Natl Acad. Sci. USA*, **82**, 4778–4782.
9. Nussbaum,R.L., Lesko,J.G., Lewis,R.A., Ledbetter,S.A. and Ledbetter,D.H. (1987) *Proc. Natl Acad. Sci. USA*, **84**, 6521–6525.
10. Straus,D. and Ausubel,F.M. (1990) *Proc. Natl Acad. Sci. USA*, **87**, 1889–1893.
11. Lisitsyn,N., Lisitsyn,N. and Wigler,M. (1993) *Science*, **259**, 946–951.
12. Liang,P. and Pardee, A.B. (1992) *Science*, **257**, 967–971.
13. Welsh,J., Chada,K., Dalal,S.S., Ralph,D., Cheng,L. and McClelland,M. (1992) *Nucleic Acids Res.*, **22**, 4965–4970.
14. Diatchenko,E., Lau,Y.F.C., Campbell,A.P., Chenchik,A., Moqadam,F., Huang,B., Lukyanov,S., Lukyanov,K., Gurskaya,N., Sverdlov,E.D. and Siebert,P.D. (1996) *Proc. Natl Acad. Sci. USA*, **93**, 6025–6030.
15. Krill,D., Shuman,M., Thompson,M.T., Becich,M.J. and Strom,S.C. (1997) *Urology*, **49**, 981–988.
16. Li,J., Yen,C., Liaw,D., Podsypanina,K., Bose,S., Wang,S.I., Puc,J., Miliaresis,C., Rodgers,L., McCombie,R., Bigner,S.H., Giovanella,B.C., Ittmann,M., Tycko,B., Hibshoosh,H., Wigler,M.H. and Parsons,R. (1997) *Science*, **275**, 1943–1947.
17. Oliner,J.D., Kinzler,K.W., Meltzer,P.S., George,D.L. and Vogelstein,B. (1992) *Nature*, **358**, 80–83.
18. Ciaparrone,M., Yamamoto,H., Yao,Y., Sgambato,A., Cattoretti,G., Tomita,N., Monden,T., Rotterdam,H. and Weinstein,I.B. (1998) *Cancer Res.*, **58**, 114–122.
19. Galaktionov,K. and Beach,D. (1991) *Cell*, **67**, 1181–1194.
20. Hubank,M. and Schatz,D.G. (1994) *Nucleic Acids Res.*, **22**, 5640–5648.
21. Velculescu,V.E., Zhang,L., Vogelstein,B. and Kinzler,K.W. (1995) *Science*, **270**, 484–487.