2598–2602 *Nucleic Acids Research, 1997, Vol. 25, No. 13*

A high throughput screening for rarely transcribed differentially expressed genes

Oliver D. von Stein, Wolf-Gerolf Thies[1] and Martin Hofmann*

Institut für Genetik (IGEN) and [1]Institute für Toxikologie (ITOX), Forschungszentrum Karlsruhe, PO Box 3640, 76021 Karlsruhe, Germany

Received March 14, 1997; Revised and Accepted May 12, 1997

ABSTRACT

A novel method combining elements of suppression subtractive hybridization with high throughput differential screening permits the efficient and rapid cloning of rarely transcribed differentially expressed genes. The experimental strategy virtually excludes the possibility of isolating false positive clones. The potential of the method is demonstrated by the isolation of 625 differentially expressed cDNAs from the metastatic adenocarcinoma cell line Bsp73-ASML when subtracted from its non-metastatic counterpart Bsp73-1AS. Northern analysis of 72 randomly selected clones demonstrated that 68 were differentially expressed with respect to Bsp73-ASML, indicating a true positive rate of 94%. Additionally, a large proportion of these clones represented rare transcripts as determined by the exposure time required to detect a signal. Sequence data indicated that of the 625 clones obtained, 92 clones scored perfect or near perfect matches with already known genes. Two hundred and eighty one clones scored between 60 and 95% homology to known human and mouse genes, whereas 252 clones scored no match with any sequences in the public databases. The method we describe is ideally suited whenever subtle changes in gene expression profiles need to be determined.

INTRODUCTION

In almost any area of biology or medicine questions arise that make it desirable to know as many possible differences in gene expression between two types of cells or between two conditions.

There exists a large repertoire of techniques that aim at producing an inventory of differential transcripts between two populations of mRNAs. Identification and isolation of differentially expressed transcripts is generally achieved by one of the following methods: differential display and related techniques (1,2); representational difference analysis (RDA) (3); enzymatic degradation subtraction (4); linker capture subtraction (5); techniques involving physical removal of common sequences (6,7). Despite the fact that all these methods have proven successful in isolation of differentially expressed genes, they all possess some specific intrinsic drawbacks. For instance, differential display restricts the analysis to differences at the 3′-end of cDNAs, so that differences in the 5′-portion of cDNAs (e.g. variants of alternatively spliced genes) are not detected. Additionally, variable reproducibility of the differential band patterns and the significant incidence of false positives make it difficult to isolate rare transcripts that are differentially expressed. Another common feature of the methods mentioned above also represents an obstacle to isolation of rare transcripts: the disproportion of concentrations of differentially expressed genes is maintained in the subtraction. RDA requires multiple rounds of subtraction, as the method fails to take into account the large differences in relative abundance of individual mRNA transcripts.

The ideal system for subtractive cloning would generate an equalized representation of differentially expressed genes irrespective of their relative abundance, would permit the monitoring of subtraction efficiency prior to the time consuming screening work and would minimize, if not completely exclude, the isolation of false positive clones.

Recently, a novel technique called subtraction suppression hybridization (SSH; 8) has been described that combines a high subtraction efficiency with an equalized representation of differentially expressed sequences: The method is based on a specific form of PCR that permits exponential amplification of cDNAs which differ in abundance, whereas amplification of sequences of identical abundance in two populations are suppressed. We have developed a system that combines the advantages of SSH with high throughput differential screening. The system allows rapid identification of differentially expressed genes and virtually excludes isolation of false positive and false negatives clones.

MATERIALS AND METHODS

Cell culture

The rat pancreatic adenocarcinoma cell lines Bsp73-1AS, Bsp73-10AS and Bsp73-ASML were cultured as described (9).

Isolation of poly(A)$^+$ RNA and cDNA synthesis

Polyadenylated RNA was isolated as previously described (10). Double-stranded cDNA was synthesized using Superscript reverse transcriptase (Gibco BRL) as follows. Aliquots of 2 µg poly(A)$^+$ RNA with 500 ng oligo(dT$_{30}$) primer in a volume of 11 µl was heated to 70°C for 10 min in a thermal cycler (Perkin

*To whom correspondence should be addressed. Tel: +49 7247 822781; Fax: +49 7247 823354; Email: hofmann@igen.fzk.d400.de

Elmer 2400) and rapidly chilled on ice. The reaction mixture was made up to 20 µl by adding 4 µl 5× first strand reaction buffer (provided with the reverse transcriptase), 2 µl 0.1 M DTT and 1 µl dNTP mix (10 mM each dATP, dGTP, dCTP and dTTP). Reverse transcription was started by adding 2 µl Superscript reverse transcriptase. From this reaction mixture, 1 µl was immediately taken out and added to a labelling mixture to monitor incorporation (tracer reaction; see below). Both reaction mixtures were incubated at 42°C for 1 h. The tracer reaction was then stopped by adding EDTA to a final concentration of 20 mM.

Tracer reaction. The efficiency of the first strand reaction was monitored by adding 1 µl of the first strand synthesis to 1 µl of a mixture containing 0.3 µCi [^{32}P]dCTP. Specific incorporation of dCTP into high molecular weight nucleic acid was determined by the TCA precipitation procedure (11). Typically ~28% of the label was incorporated in the first strand reaction.

Following first strand synthesis, second strand synthesis was performed by adding 91.8 µl sterile, bidistilled water, 32 µl 5× second strand buffer [94 mM Tris–HCl, pH 6.9, 453 mM KCl, 23 mM $MgCl_2$, 750 µM β-NAD and 50 mM $(NH_4)_2SO_4$], 3 µl 10 mM mixed dNTPs stock, 6 µl 0.1 M DTT, 2 µl *Escherichia coli* DNA ligase (7.5 U/µl), 4 µl *E.coli* DNA polymerase I (10 U/µl) and 0.7 µl *E.coli* RNase H (2 U/µl). The contents were mixed and 5 µl were removed for the radioactive incorporation assay (as above). Both reactions were incubated at 16°C for 2.5 h. The double-stranded cDNA in the unlabelled reaction was blunted by the addition of T4 polymerase, followed by further incubation at 16°C for 20 min before both reactions were stopped by addition of EDTA. As before, labelled second strand cDNA synthesis was monitored by TCA precipitation and by resolving the product on an alkaline agarose gel (11).

Generation of a subtracted library by SSH

SSH was performed between Bsp73-1AS ('driver') and Bsp73-ASML ('tester') using the PCR-Select™ cDNA Subtraction Kit (Clontech) according to the manufacturer's recommendations, except for modifications of the PCR and hybridization conditions. All PCR and hybridization steps were performed on a Perkin-Elmer 2400 thermal cycler.

For the first hybridization the mixture of driver and tester cDNAs was denatured at 100°C for 20 s and then cooled over 1 min to 68°C and maintained at that temperature for ~8 h. For the second hybridization, driver cDNA was denatured at 100°C for 20 s then added directly to the pooled mix of the two previous hybridizations and allowed to incubate at 68°C for 20 h. It was necessary to alter the PCR conditions (see below) such that amplification of unwanted sequences was kept to a minimum. All other procedures for generation of the subtracted library were done according to the manual provided with the cDNA subtraction kit (Clontech).

Evaluation of subtraction efficiency

Double-stranded cDNA from the tester (Bsp73-ASML) and the driver (Bsp73-1AS) were separately digested with *Rsa*I (a four base cutter, as used for the construction of the initial subtracted library). *Hin*dIII linkers were added to the tester cDNA and *Eco*RI linkers to the driver cDNA (*Hin*dIII linker, 5′-ATCGTCAAGCTTCA-AGTTAGCATCG-3′, 5′-GCTAACTTGAAGCTTGACGAT-3′; *Eco*RI linker, 5′-TAGTCCGAATTCAAGCAAGAGCACA-3′, 5′-CTCTTGCTTGAATTCGGACTA-3′). Free linkers were removed by preparative agarose gel electrophoresis and the cDNA mixture was amplified as follows. An aliquot of 1 µl adaptor-ligated cDNA was diluted into 1 ml H_2O and amplified in a 50 µl reaction using the appropriate primers (*Hin*dIII and *Eco*RI primers), a standard PCR buffer (Pharmacia), 200 µM dNTPs and 2 U *Taq* polymerase (Pharmacia). Cycling parameters were as follows: 30 cycles of 94°C for 20 s, 52°C for 20 s and 72°C for 2 min. For the subtracted cDNA, PCR conditions were as follows: 27 cycles each of 94°C for 30 s, 68°C for 30 s and 72°C for 2 min. Only the subtracted cDNA was subjected to a second round of PCR (nested), using the same PCR conditions with the exception that 12 cycles were performed. Equal amounts of amplified cDNA from the driver, the tester and of the subtracted library were resolved on a 1.5% agarose gel, blotted and transferred onto nylon membrane. Hybridizations were carried out under stringent conditions in 0.5 M Na_2PO_4, pH 7.2, 7% SDS at 65°C (12). Filters were washed twice in 2× SSC, 0.5% SDS at 68°C, then once in 0.1× SSC, 0.1% SDS at 68°C.

Cloning into a TA vector

After evaluation of the subtraction efficiency the subtracted library cDNA was cloned directly into pCRII.1 (TA Cloning Kit, Invitrogen). Prior to ligation the subtracted PCR cDNA mix was incubated for a further 1 h at 72°C with additional dATP and *Taq* DNA polymerase (Pharmacia) to ensure that most of the cDNA fragments contained 'A overhangs'. Approximately 100 ng PCR-amplified cDNA were ligated without further purification into 50 ng vector and the ligation mixture was introduced into Electromax bacterial strain DH10B (Gibco BRL) by electroporation (1.8 kV) using an *E.coli* pulser (BioRad). The library was plated onto 22 × 22 cm ampicillin-containing agar plates and bacteria were grown until colonies were visible. Bacteria were then washed off in LB medium, aliquoted and frozen in 10% DMSO. For library screening the titre was determined and bacteria were plated onto 22 × 22 cm agar plates containing 100 µg/ml ampicillin, 100 µM IPTG and 50 µg/ml X-Gal. Plates were incubated at 37°C until small colonies were visible then incubated further at 4°C until blue/white staining could be clearly distinguished.

Reverse Northern high density blots and screening

A total of 5000 individual recombinant clones were picked and used to inoculate 52 sterile 96-well microtitre plates containing LB medium and ampicillin at 100 µg/ml. After incubation of bacteria on a gyratory shaker for 4 h at room temperature, 5 µl bacterial culture were transferred into 15 µl sterile water in PCR tubes. (This part of the protocol was done in a 96 tube format using a Perkin-Elmer 9600 thermal cycler. Pipetting of PCR mixes was done using a multichannel pipette.) The bacteria were lysed by heating to 100°C for 5 min. Samples of 5 µl bacterial lysate were used to PCR amplify cloned inserts in 50 µl reactions using standard PCR buffer (Pharmacia), 200 µM dNTPs, 2 U *Taq* polymerase (Pharmacia) and 10 µmol M13 rev and M13(–20) primers (which flank the multiple cloning site of pCRII.1) under the following conditions: 30 cycles each of 94°C for 20 s, 48°C for 20 s and 72°C for 45 s.

After amplification, 12 µl were loaded onto high density gels (Centipede™ gel electrophoresis chambers; Owl Scientific, Woburn, MA). PCR products were denatured and alkaline blotted

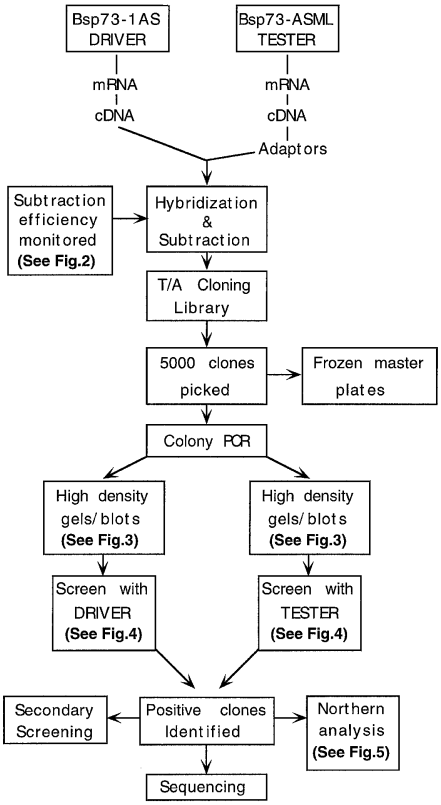

Figure 1. A schematic representation of the method.

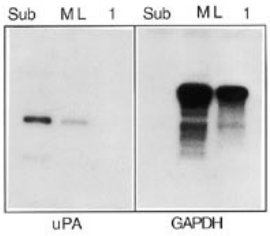

Figure 2. Equal amounts of PCR-amplified driver, tester and subtracted cDNA were fractionated on a 1.4% agarose gel, blotted and hybridized with [^{32}P]dCTP-labelled uPA (left) and GAPDH (right). Lane 1, Bsp73-1AS; lane ML, Bsp73-ASML; lane Sub, subtracted cDNA.

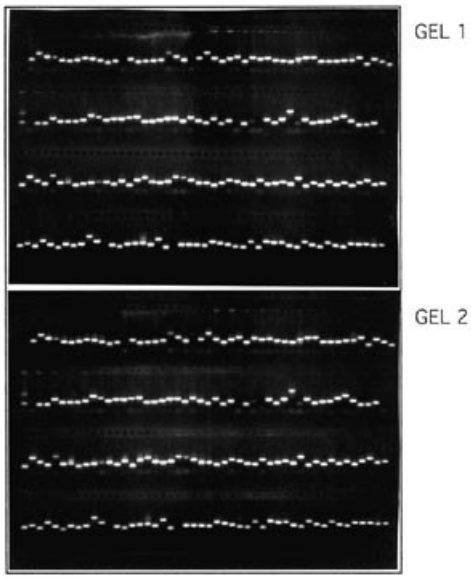

Figure 3. Colony PCR was performed as described and the products resolved on 1% agarose gels in parallel (GEL 1 and GEL 2). The gels were then stained with ethidium bromide and photographed to ensure equal loading.

onto nylon membranes. The filters were hybridized under stringent conditions (7% SDS in 0.5 M $NaPO_4$, pH 7.2; 12) with equivalent amounts of ^{32}P-labelled double-stranded cDNA of approximately equal specific activity derived from driver and tester mRNA respectively. Filters were washed under stringent conditions (see above) and exposed to phosphorimager plates overnight (equivalent to 8 days exposure to conventional film). In addition, the filters were exposed to conventional film for up to 12 days at –80°C and the signals of like clones compared.

Northern blotting

Aliquots of 4 µg poly(A)$^+$ RNA were fractionated on a 1.4% formaldehyde–agarose gel and blotted in 10× SSC overnight onto Hybond N$^+$ membrane (Amersham). Filters were hybridized at 65°C with the probes using QuickHyb™ (Stratagene) and washed according to the manufacturer's guidelines. Probes were generated by PCR amplification of the insert of interest, gel purifying and ^{32}P-labelling the cDNA (ReadyPrime; Amersham). Unincorporated label was removed prior to hybridization using an Elutip (Schleicher & Schüll) according to the manufacturer's specifications. Filters were then exposed to autoradiography film at –80°C for from 4 h to >4 days.

RESULTS AND DISCUSSION

Phenotypic differences between cells of the same organism are determined by differential expression of identical genes. Identification of the genes responsible for these phenotypic differences requires methods which rapidly and efficiently compare the transcripts expressed in the two cell types and which permit isolation of those transcripts found in only one cell type. The recent description of a novel equalizing cDNA subtraction method, termed suppression subtractive hybridization (SSH; 8), provides the technical basis for such a comparison. In an attempt to identify changes in gene expression patterns that accompany the progression from a non-metastatic to a metastatic tumour, we subtracted the highly metastatic adenocarcinoma cell line cell line Bsp73-ASML from its non-metastatic counterpart Bsp73-1AS. Both cell lines have previously been used for isolation of a metastasis-associated gene, CD44v (13).

The flow diagram of the resulting screening system is depicted in Figure 1. It starts with the two populations of mRNAs that are to be subtracted. The 'driver' is defined as the population of mRNAs that will be eliminated during subtraction, whereas the 'tester' population of mRNAs contains in addition the differentially expressed genes of interest. Both mRNA populations are reverse transcribed and subjected to the SSH procedure as described by Diatchenko *et al.* (8). One important feature of the method is that the efficiency of the subtraction process can be easily monitored

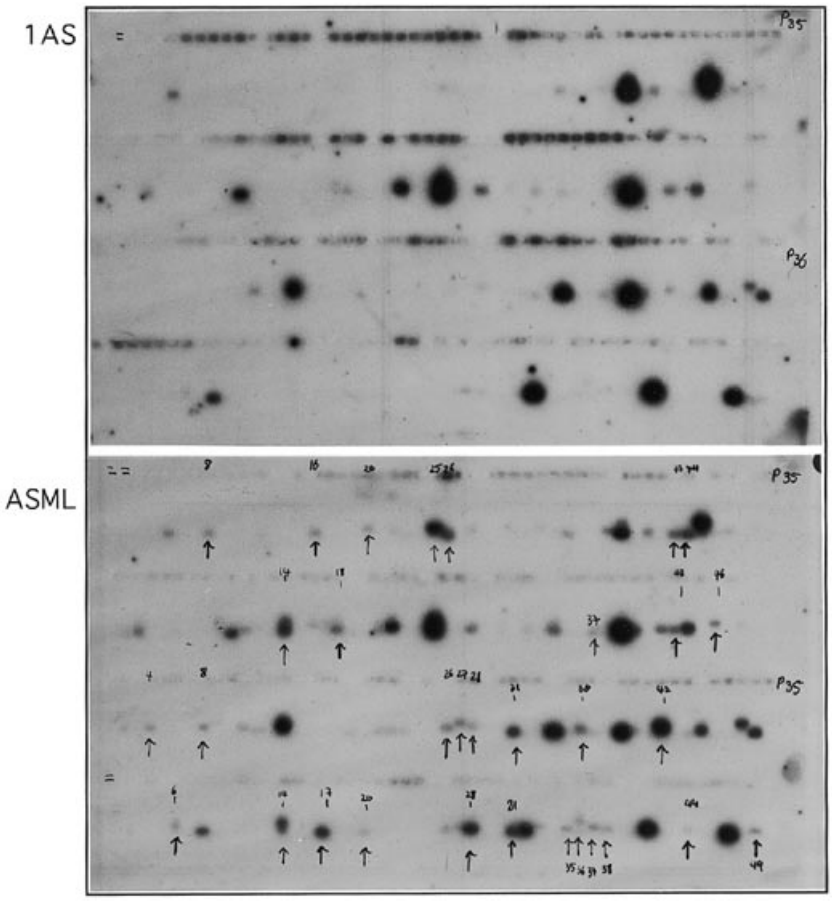

Figure 4. Duplicate filters were hybridized with double-stranded ^{32}P-labelled 1AS and ASML cDNA of equal specific activity under stringent conditions. After washing, the filters were exposed to autoradiography film at –80°C for 10–12 days. The arrows indicate those clones that show differential hybridization and therefore represent clones harbouring cDNA fragments that originate from the ASML cell line only. Additionally, clones that show only very weak differential expression are also detected.

before any further processing of the subtracted cDNAs is undertaken. The degree of subtraction efficiency can be determined by either monitoring depletion of transcripts common to both populations or enrichment of sequences specific to one population after subtraction.

From previous studies we knew that the urokinase plasminogen activator (uPA) gene is expressed in Bsp73-ASML cells only and thereby fulfils the criterion of a positive control for the subtraction. Southern blot hybridization analysis with a uPA probe shows enrichment of the uPA cDNA in the subtracted cDNA population (Fig. 2, left). Expression of genes existing in both populations (e.g. housekeeping genes) should be drastically reduced. Glyceraldehyde phosphate dehydrogenase (GAPDH) in this context functions as a negative control for the subtraction. As can be seen in Figure 2 (right), GAPDH cDNA is not detectable, indicating very efficient removal. These two hybridizations demonstrate just two simple ways to monitor subtraction efficiency. Further possible controls include hybridization of Northern blots of tester and driver mRNA with a radiolabelled probe of the complex population of selected cDNAs enriched by SSH (see also 8).

Once the subtraction efficiency is shown to be satisfactory the subtracted cDNA is cloned into a vector containing 'T overhangs' (e.g. the Invitrogen TA Cloning Kit). The library is transformed into bacteria by electroporation and plated onto agar plates.

Initial attempts to identify rarely transcribed differentially expressed genes by conventional differential screening (colony filter lifts and hybridization) failed due to non-specific binding to bacterial nucleic acids and proteins by the complex screening probe (*Rsa*I-restricted double-stranded cDNA of the tester or driver respectively). This made it extremely difficult to detect clones harbouring rarely transcribed genes against the non-specific background. To overcome this problem, the library was plated on agar plates containing ampicillin, X-Gal and IPTG. Recombinant (white) clones were picked and used to inoculate LB medium in 96-well microtitre plates. Bacteria were allowed to grow at room temperature for 4 h before aliquots of bacterial suspension were taken and heat denatured as described in Materials and Methods. Cloned inserts were amplified by colony PCR and the PCR products resolved in duplicate on special high density gels (200 clones/gel, Centipede™ chamber; Owl Scientific). Care was taken to ensure equal loading of duplicate gels to permit direct comparison of hybridization signal intensities (Fig. 3). Following immobilization of the nucleic acids on nylon membranes, duplicate filters were hybridized with either radiolabelled, restricted, double-stranded cDNA from tester or driver. Restriction of the double-stranded cDNA probes is of importance, as we found that the resulting hybridization signals were much more pronounced as against those obtained with undigested probes.

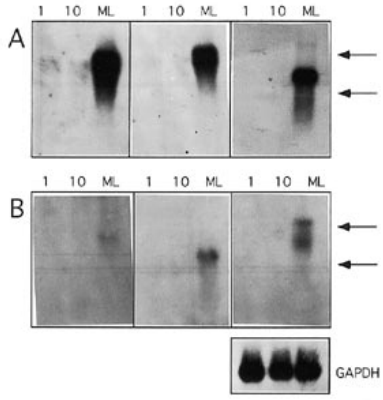

Figure 5. Three clones showing strong differential signals and three clones showing weak differential signals identified by the initial screening were hybridized against 4 µg poly(A)$^+$ RNA derived from the cell lines indicated. Lane 1, Bsp73-1AS; lane 10, Bsp73-10AS; lane ML, Bsp73-ASML. Exposure times ranged from 4–12 h (**A**) to >4 days (**B**) at –80°C. The top and bottom arrows on the right of the figure indicate the positions of 28S and 18S RNA respectively.

Filters were exposed to conventional film for up to 12 days and the signals of identical clones were compared.

Faint signals representing rarely transcribed differentially expressed genes could be distinguished (Fig. 4). This so-called 'inverted Northern' hybridization allows rapid and efficient identification of clones harbouring differentially expressed cDNAs. The full range from abundant to rarely transcribed differentially expressed genes can be isolated by this procedure, as is evident from Northern hybridizations that were performed to confirm differential expression. Some clones representing abundant transcripts required 4–12 h exposure (Fig. 5A), whereas clones representing low abundance transcripts required more than 4 days at –80°C in order to detect a signal on conventional film (Fig. 5B).

Positive clones were grouped according to their hybridization signal intensity (strong signals 114, middle intensity signals 207 and weak signals 304) and used to inoculate new 96-well master plates in duplicate, of which one was used for further analysis (Northern analysis/sequencing), whereas the other was frozen in 10% DMSO at –80°C as a reference stock.

Four of the clones identified by the procedure outlined above gave a false positive signal (so far ~70 clones have been analyzed and confirmed by Northern hybridization). This equates to a true positive rate of ~94% and therefore confirmation of differential expression by Northern analysis for each clone obtained by our procedure is probably unnecessary. Obviously positive signals seen with reverse Northern blots represent truly differentially expressed cDNA clones. Automated sequencing proves advantageous in rapidly analysing the complete library and thus providing data on both identity and redundancy of the cDNA clones obtained.

In our initial screening using the method described here we analyzed 5000 clones. Approximately 72% of them showed a hybridization signal with radiolabelled cDNA derived from the tester and 50% with labelled cDNA from the driver. Clones that failed to hybridize may represent extremely rare transcripts and thereby lie below the sensitivity limit or 'reverse Northerns'. Additionally, some 10–12% of the total number of clones picked for screening did not contain an insert as determined by restriction analysis of 50 randomly picked clones. A total of 625 clones showed differential hybridization, which equates to 12.5% of all clones in the subtracted library harbouring cDNAs specific for the tester cell line Bsp73-ASML.

Sequence analysis of the 625 clones obtained indicated that 92 clones scored perfect or near perfect matches with already known genes. Two hundred and eighty one clones scored homologies ranging between 60 and 95% with known human and mouse genes, whereas 252 clones scored no match with any sequences in the public databases. The sequence data also indicated that the majority of the clones were picked two to six times, indicating a small degree of redundancy within the subtracted library. We estimate that genes which differ dramatically in their abundance between two populations should be identified within the first 100 picked clones from the initial library. Isolation of genes representing rare abundance transcripts requires the analysis of 500 or more picked clones from the initial library.

In contrast to display techniques, the method described here allows identification of hundreds of differentially expressed genes in one hybridization experiment. After identification of clones representing differentially expressed cDNAs (in our case between the two cell lines Bsp73-1AS and Bsp73-ASML), criteria other than differential expression can be applied in further screenings. For example, filters containing inserts of previously identified differentially expressed clones can be hybridized to radiolabelled cDNA from other sources, thereby providing insight concerning their absence or presence in normal and/or tumour tissues respectively. We have shown that SSH used in conjunction with high throughput differential screening allows rapid and easy identification of rarely and frequently transcribed, differentially expressed genes. The system introduced in this paper virtually excludes the possibility of isolating false positive clones and thereby drastically reduces the workload on the experimenter.

ACKNOWLEDGEMENTS

This work was supported by a grant of the Dr Mildred Scheel Stiftung für Krebsforschung (W53/94/Ho1). The authors would like to thank Harald König, Jonathan Sleeman and Klaus Störmann for valuable discussions and hints and Conny Leicht for her excellent technical assistance.

REFERENCES

1 Liang,P. and Pardee,A. (1992) *Science*, **257**, 967–970.
2 Sokolov,B.P. and Prockop,D.J. (1994) *Nucleic Acids Res.*, **22**, 4009–4015.
3 Lisitsyn,N., Lisitsyn,N and Wigler,M. (1993) *Science*, **259**, 946–951.
4 Zeng,J., Gorski,R.A. and Hamer,D. (1994) *Nucleic Acids Res.*, **22**, 4381–4385
5 Yang,M. and Sytkowski,A.J. (1996) *Anal. Biochem.*, **237**, 109–114.
6 Akopian,A.N. and Wood,J.N. (1995) *J. Biol. Chem.*, **270**, 21264–21270.
7 Deleersnijder,W., Hong,G., Cortvrindt,R., Poirier,C., Tylzanowski,P., Pittois,K., Van Marck,E. and Merregaert,J. (1996) *J. Biol. Chem.*, **271**, 19475–19482.
8 Diatchenko,L., 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.
9 Matzku,S., Komitowski,D., Mildenberger,M. and Zöller,M. (1983) *Inv. Met.*, **3**, 109–123.
10 Rahmsdorf,H.J., Schönthal,A., Angel,P., Litfin,M., Rüther,U. and Herrlich,P. (1986) *Nucleic Acids Res.*, **15**, 1643–1659.
11 Ausubel,F.M., Brent,R., Kingston,R.E., Moore,D.D., Seidman,J.A., Struhl,K. (1987) *Current Protocols in Molecular Biology.* John Wiley & Sons Inc., USA.
12 Church,G.M. and Gilbert,W. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 1991–1995.
13 Günthert,U., Hofmann,M., Rudy,W., Reber,S., Zöller,M., Haußmann,I., Matzku,S., Wenzel,A., Ponta,H and Herrlich,P. (1991) *Cell*, **65**, 13–25.