Nucleic Acids Research, 2005, Vol. 33, No. 16 ***e136***
doi:10.1093/nar/gni136

A novel transcriptome subtraction method for the detection of differentially expressed genes in highly complex eukaryotes

L. Li*, D. Techel[1], N. Gretz and A. Hildebrandt[2]

Centre for Medical Research (ZMF), University of Heidelberg, Germany, [1]Institute for Legal Medicine, Bremen, Germany and [2]Department of Environmental Research and Technology, University of Bremen, Germany

Received April 22, 2005; Revised July 4, 2005; Accepted August 19, 2005

ABSTRACT

We have designed a novel transcriptome subtraction method for the genome-scale analysis of differential gene expression in highly complex eukaryotes, in which suppression subtractive hybridization (SSH) is performed first to enrich the target and, after exchange of adapters, negative subtraction chain (NSC) is then used to eliminate the remaining background. NSC evolved from differential subtraction chain (DSC). We designed novel adapters which make the subtraction system more robust. SSH and NSC were then combined to successfully detect differentially expressed genes in *Solanum*. The combined technique improves qualitatively upon SSH, the only commercially available transcriptome subtraction system, by detecting target genes in the middle abundance class, to which most differentially expressed genes in highly complex eukaryotes are expected to belong. The main advantage of the combined technique with SSH/NSC is its ability to isolate differentially expressed genes quickly and cost-efficiently from non-standard models, for those microarrays are unavailable.

INTRODUCTION

There are a few strategies available for the analysis of differential gene expression on a whole genome scale, but all of them with intrinsic disadvantages. Differential display (1) is suitable only for less complex genomes. Microarray techniques (2) are expensive and with limited use only for standardized models. Much time is required to set up a gene expression profile for an eukaryote with serial analysis of gene expression (SAGE) (3). Transcriptome subtraction promises, at least theoretically, to be a fast and cheap way to analyze differential gene expression, also of non-standard models, on a whole genome scale.

There are two methods frequently used for transcriptome subtraction: representation differential analysis (RDA) (4) and suppression subtractive hybidization (SSH) (5). RDA requires very intensive hands-on work and is susceptible to contamination. SSH enriches target sequences dramatically, but only one step of subtraction is possible, resulting in an increase of background sequences in the final product, as the genome complexity increases. For example, when two eukaryotic genomes are compared with SSH, the background sequences in the final SSH-product often get so numerous that the target sequences are totally masked. Another patented system called differential subtraction chain (DSC) (6), whose authors stated it is robust and easy to use, was not reproducible in our hands.

In this study, we show that DSC might have some weakness because of its adapters. We re-designed the adapters to make the subtraction system more robust and developed an improved technique called negative subtraction chain (NSC). The sensitivity of NSC is expected to be limited, because low abundance sequences might be degraded during the subtraction by mung bean nuclease, whose activity is essential for the function of this assay. NSC was therefore combined with SSH to detect differentially expressed genes in highly complex eukaryotes. With its higher sensitivity, the combined technique shows a significantly improved resolution compared with SSH alone, the only commercially available method for transcriptome subtraction.

MATERIALS AND METHODS

Applying DSC and NSC

We attempted to repeat a DSC experiment, in which a Lambda-fragment attached to a tester-adapter was reported to be completely subtracted by the same fragment equipped with a driver-adapter. Lambda DNA was cut with HindIII and the 564 bp fragment was ligated to adapters with sequences as

*To whom correspondence should be addressed. Tel: +49 621 383 3769; Fax: +49 621 383 2108; Email: li.li@gmx.de

© The Author 2005. Published by Oxford University Press. All rights reserved.

The online version of this article has been published under an open access model. Users are entitled to use, reproduce, disseminate, or display the open access version of this article for non-commercial purposes provided that: the original authorship is properly and fully attributed; the Journal and Oxford University Press are attributed as the original place of publication with the correct citation details given; if an article is subsequently reproduced or disseminated not in its entirety but only in part or as a derivative work this must be clearly indicated. For commercial re-use, please contact journals.permissions@oxfordjournals.org

described. The products were amplified by PCR and the driver-adapter was then removed with HindIII. Five nanogram tester and 500 ng driver were co-purified in 32 µl 3× EE buffer and covered with oil. Eight µl of 5 M NaCl was added after denaturation of the DNA at 100°C for 5 min, followed by a hybridization step at 67°C for 23 h. The DNA was then purified and treated with 10 U of mung bean nuclease at 30°C for 30 min in a total volume of 50 µl, before an aliquot was taken to monitor the subtraction efficiency by PCR (95°C for 1 min, then 35 cycles of 95°C for 15 s and 68°C for 3 min). The remaining DNA was purified for the next hybridization step. The hybridization and the treatment with mung bean nuclease were repeated for another two rounds.

NSC and DSC were performed in parallel. The procedure of NSC is the same as described for DSC, except that novel adapters/primers were applied, and the driver-adapter was not removed before the hybridization. The sequences of adapters/primers applied are listed below, with complementary sequences underlined and the common sequences between NSC adapters shown in bold.

The tester-adapter/primer for DSC: 5′-(d)AGCACTCTCCAGCCTCTCACCGCA-3′/5′-(d)AGCTTGCGGTGA-3′

The driver-adapter/primer for DSC: 5′-(d)ACCGACGTCGACTATCCATGAACA-3′/5′-(d)AGCTTGTTCATG-3′

The tester-adapter for NSC: 5′-(d)CAGTCAGAGAGCTCTCACA-3′/5′-(d)AGCTTGTGAGAG-3′

The driver-adapter for NSC: 5′-(d)GACACTCTCACCTCTCACA-3′/5′-(d)AGCTTGTGAGAG-3′

The tester-primer for NSC: 5′-(d)CAGTCAGAGAG**CTCTCACAAGCT**-3′

The driver-primer for NSC: 5′-(d)GACACTCTCAC**CTCTCACAAGCT**-3′.

To check their reproducibility, the application of DSC and NSC on the lambda fragment was repeated two more times. To check the robustness of these systems, we applied them also on other targets, including the 603 bp fragment from PhiX17 DNA/HaeIII marker and a 244 bp fragment from an *E.coli* gene (which was also used to make artificial testers as described below). The adapters/primers applied were slightly modified to contain a suitable restriction site on the ligation end.

Detection of differentially expressed genes in *Solanum* using SSH/NSC

Leaves were harvested from one month-old potato plants. RNA was isolated as described by Chomczynski and Sacchi (7) and mRNA purified with oligo (dT)-cellulose (USB Corp.) according to the instructions of the manufacturer. CDNAs were synthesized using the template switch technique (Smart^{TM} PCR cDNA synthesis kit; Clontech). For the transcriptome subtraction, the target was first enriched by SSH, and after exchange of the adapters, the remaining background was eliminated by NSC. Since cDNAs digested with Mbo1 were used for the further subtraction with NSC, all adapters/primers applied were slightly modified to contain a restriction site for Mbo1 on the ligation end.

Two SSH assays, referred as $SSH_{1/2}$ and $SSH_{2/2}$, were first performed until the first PCR step as described (Clontech, PCR-Select cDNA Subtraction Kit). $SSH_{1/2}$ used cDNA of plant 1 as tester and that of plant 2 as driver. $SSH_{2/2}$ used cDNA of plant 2 as tester and driver. A 20 µl $SSH_{1/2}$ product was then digested with Mbo1 to remove its adapter, purified and dissolved again in 20 µl of H_2O, of which 5 µl were used for the ligation to 100 pmol of the tester-adapter for NSC in a total volume of 20 µl. A 4 µl sample of this ligated product was used as a template for the synthesis of NSC tester in a 25 µl PCR reaction. The same PCR profile as that for the second PCR of SSH was applied, with a pre-incubation step at 72°C for 5 min added. This high temperature denatures the double strand of the adapters and therefore separates the unligated oligonucleotides from the ligated structure. Recombinant Taq polymerase (Invitrogen) was added 2 min after the beginning of the incubation to fill in the sticky ends of the adapters. NSC driver (125 µl) was synthesized in a similar way after the driver-adapter was ligated to the product of $SSH_{2/2}$.

One microlitre tester (50 ng) and 100 µl driver (5 µg) for NSC were purified together in 2 µl NSC buffer (10 mM EPPS, pH 8; 1 mM EDTA; 0.5 M NaCl) and covered with mineral oil. After denaturation at 99°C for 5 min, the hybridization was carried out at 67°C for 10–12 h. The reannealed products were purified and then treated with 5 U of mung bean nuclease at 37°C for 30 min in a total volume of 20 µl. One microlitre of the treated DNA was diluted with 40 µl of H_2O to serve as a template for the PCR (35 cycles of 95°C for 30 s, 68°C for 30 s and 72°C for 90 s) to examine the subtraction efficiency and the remainder was purified for the next hybridization. The hybridization and the treatment with mung bean nuclease steps were repeated for another two rounds.

The specificity of the reamplified subtracted product was verified by Southern blot, in which the product was labelled with Digoxigenin-11-dUTP (1 ng/ml) by random priming and hybridized to the original tester and driver cDNAs (1 µg each) in parallel. The subtraction product was also cloned with the TOPO TA Cloning Kit for Sequencing (Invitrogen). Fifteen randomly chosen clones were then processed for the isolation of plasmids. One nanogram of each plasmid were dotted to nylon membranes and hybridized to the original tester or driver cDNAs labelled with Digoxigenin-11-dUTP. Samples of 50 pg of unlabelled tester and driver cDNA were also blotted to the nylon membranes to serve as positive controls for the hybridization. The inserts were then characterized by sequencing, with Sp6 as a sequencing primer.

Evaluating the sensitivity of the combined technique

The sensitivity of the combined technique was evaluated with a DNA fragment from the *E.coli* genome as an artificial target, which is not present in *Solanum*. It was amplified by PCR with the primers 5′-CGACGCTCACACCGATACCATC-3′ and 5′-CCATACCTGTTCACCGACGACG-3′. The resulting PCR product containing two Mbo1 restriction sites was cut by Mob1 and separated in a 0.8% agarose gel. The 244 bp middle part between the two Mbo1 sites was isolated from the gel and quantified with a photometer. Samples of 0, 2, 10 or 50 pg of this fragment were spotted to each microgram of *Solanum* cDNAs to generate artificial testers. We assume the cDNA of a typical plant gene is between 2 and 3 kb and, when it is digested with Mob1, generates 10 different target sequences in average. Therefore, these artificial testers with 0, 2, 10 and 50 p.p.m. of the artificial target fragment correspond to target gene concentrations of 0, 20, 100 and 500 p.p.m.

The subtraction with SSH and NSC was performed as described above.

RESULTS

Development of NSC based on DSC

DSC is intended to convert amplifiable tester sequences into non-amplifiable driver sequences by repeated hybridization of testers and drivers that are attached to different adapters and the degradation of the tester-adapter in hetero-hybrids. The result was reported to be a double exponential elimination of amplifiable (background) sequences in the testers which have counterparts in the drivers, while preserving the target sequences.

We attempted to repeat a DSC experiment, in which 5 ng of lambda fragment attached to the tester-adapter were reported to be completely subtracted by 500 ng of the same fragment attached to the driver-adapter in three rounds of DSC. This result was not reproducible in our hands: after DSC the tester remained intact.

We hypothesized that DSC in its published version might have some weakness because of its adapters. Degradation of the tester overhangs in hetero-hybrids is essential for efficient subtraction with DSC, but the tester overhangs in its hetero-hybrids are 24–28 bp (24 bp from the adapters and 0–4 bp from the sticky ends depending on the restriction enzyme used) long, and they will form stable double-stranded DNA with their complementary sequences from other hetero-hybrids and are therefore no longer accessible to single strand-specific mung bean nuclease. To test this hypothesis, we modified the adapter design according to the following criteria: first, the adapters/primers must be long and specific enough for PCR; second, the tester overhang in hetero-hybrids, which is identical to the non-complementary part of the tester-adapter to the driver-adapter, must be short enough to remain single-stranded during mung bean nuclease treatment; and third, the remaining part of the tester-adapter in the hetero-hybrids after mung bean nuclease treatment should be short enough to prevent the tester-primer from annealing during the PCR for the examination of the subtraction efficiency. The NSC adapters/primers were designed to fulfil these criteria: they are only 23 bp long and consist of a common sequence of 12 bp at the ligation end and a specific sequence of 11 bp. We call this subtraction method with novel adapters/primers NSC (Figure 1), in that its development was inspired by the idea of DSC.

NSC was tested in direct comparison to DSC. Although both methods in principle differ only on their adapters/primers, an efficient subtraction was achieved only by NSC, but not by DSC. In the NSC product, the tester is completely subtracted and no longer reamplifiable by PCR, whereas still intact, PCR amplifiable tester sequences exist in the DSC product. Repeated experiments on three different targets produced consistent results. A representative picture for the functionality of DSC and NSC is shown in Figure 2. During NSC the subtraction is progressive, Luo *et al*. (6) described this progress brilliantly in a mathematical model. But the effect of the progressive subtraction cannot be seen in Figure 2, because a classical qualitative, instead of a quantitative, real-time PCR was used to amplify the remaining tester.

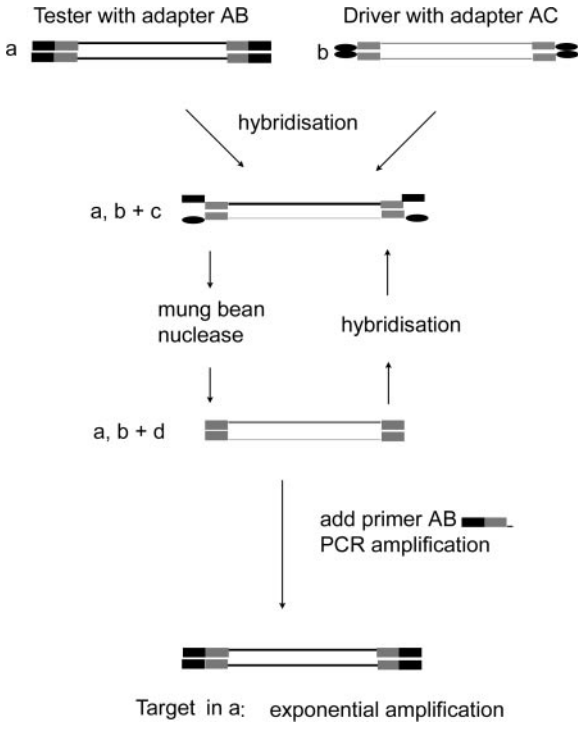

Figure 1. Flow chart of NSC. Tester and driver are attached to adapters which consist of a specific and a common sequence, then subjected to hybridization. The specific part of the tester-adapter in hetero-hybrids is short enough to remain single stranded and can be degraded by mung bean nuclease. For progressive subtraction, the steps of hybridization and mung bean nuclease treatment are repeated.

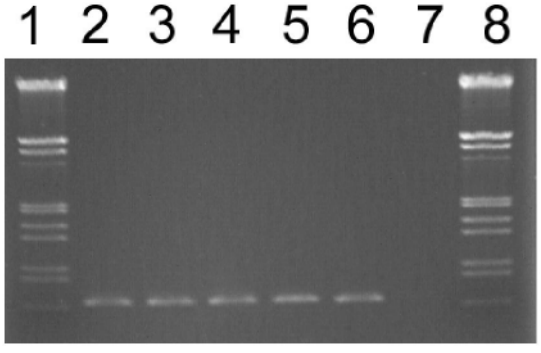

Figure 2. Applying DSC and NSC on a lambda fragment. Efficient subtraction was achieved only by NSC, but not by DSC: 5 ng Lambda/HindIII fragment attached to the tester-adapter was subtracted with 500 ng of the same fragment attached to the driver-adapter. Lanes 1 and 8, Lambda/HindIII + EcoRI marker; lanes 2–4, PCR reamplified product with the tester-primer after 1–3 rounds of DSC; lanes 5–7, PCR reamplified product with the tester-primer after 1–3 rounds of NSC.

Detection of differential gene expression in highly complex eukaryotes with SSH/NSC

The idea to combine SSH and NSC evolved after consideration of the shortcomings and advantages of both methods. On one hand, SSH enriches target sequences dramatically, but it is still not efficient enough to directly isolate differentially expressed genes from highly complex eukaryotes. On the other hand, NSC is an efficient subtractive method, but its sensitivity

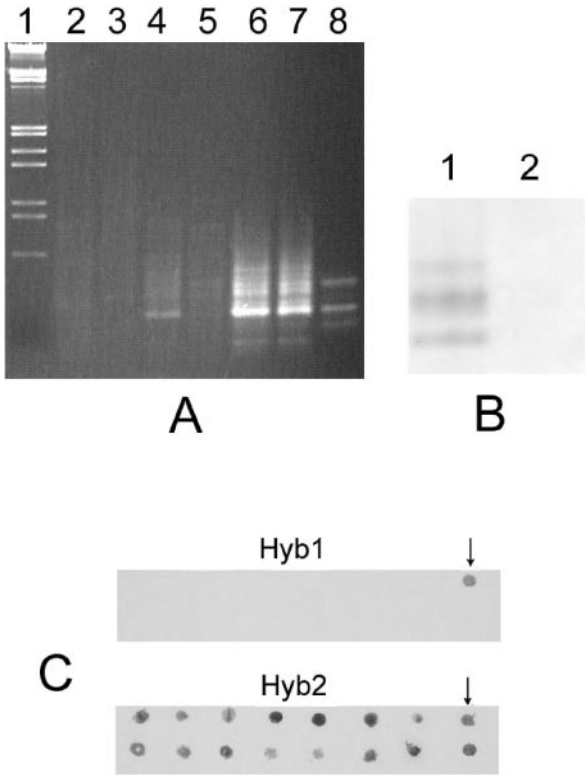

Figure 3. (**A**) Detection of differentially expressed genes in *Solanum* with SSH/NSC. Lane 1, Lambda/HindIII + EcoRI marker; lane 2, cDNA of plant 1, cut with Mbo1; lane 3, cDNA of plant 2, cut with Mbo1; lane 4, NSC tester; lane 5, NSC driver; lanes 6–8, PCR reamplified products after 1–3 rounds of NSC. (**B**) Southern blot analysis of the subtracted product. The final product after transcriptome subtraction was labelled with Digoxigenin and hybridized to the original tester (lane 1) and to the original driver cDNA (lane 2) cut by Mbo1. (**C**) Dot blot analysis of the cloned subtracted product. Plasmids of 15 randomly chosen clones were blotted on the nylon membrane and hybridized to the labelled driver cDNA (Hyb1) or the labelled tester cDNA (Hyb2). Arrows point to positive controls of the hybridization.

might not be very high. Mung bean nuclease, which is essential for the function of NSC, can also degrade double-stranded DNA, thereby causing the loss of target sequences of lower abundance. Therefore, both methods were combined for the detection of differentially expressed genes in highly complex eukaryotes, with SSH applied first to enrich the target and then NSC to further eliminate the remaining background. Because they both are equipped with different adapters, between both subtraction procedures a step of adapter exchange is required.

With *Solanum* cDNAs, the combined technique SSH/NSC was shown to be able to directly identify the differentially expressed genes in highly complex eukaryotes. When run on a gel, the subtraction products were visible as distinct DNA fragments (Figure 3A). These products were then used as a probe to hybridize on cDNAs of the tester and the driver. Distinct hybridization signals were seen only on the tester cDNA (Figure 3B).

The subtraction product was also cloned and plasmids of fifteen randomly chosen clones were dot-blotted to nylon membranes, before hybridized to the original tester or driver labelled with Digoxigenin-11-dUTP. All 15 clones were reconfirmed to be specific for the tester by dot-blot (Figure 3C). These clones were also characterized by sequencing. Four, two

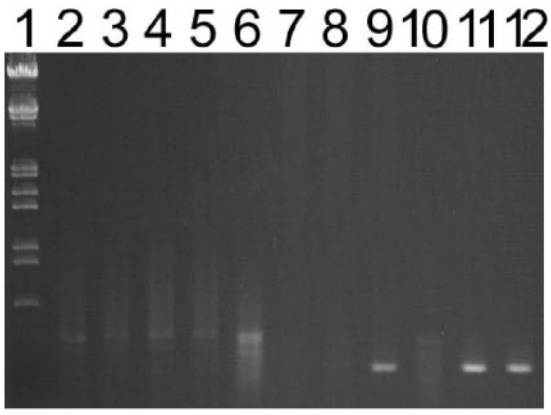

Figure 4. Sensitivity of SSH and SSH/NSC. With the combined technique, it was possible to detect an artificial target gene of 100 p.p.m., whereas with SSH only a target of ⩾500 p.p.m. was detected directly. Lane 1, Lambda/HindIII + EcoRI marker; lanes 2–5, PCR product for the control on the ligation success between the tester-adapter and testers (with 0, 20, 100 or 500 p.p.m. artificial target); lanes 6–9, SSH-products from testers with 0, 20, 100 or 500 p.p.m. artificial target; lanes 10–12, SSH/NSC-products from testers with 20, 100 or 500 p.p.m. artificial target.

and four clones turned out to carry sequences highly homological to plant proteins xyloglucan endo-transglycosylase, serine/threonine kinase and At5g13260, respectively. Two clones carried sequences coding for other putative proteins and the remaining three clones carried uncharacterized DNA sequences.

The combined technique turned out to be more sensitive than SSH alone (Figure 4). With SSH alone, the lowest target gene concentration that could be detected directly without tedious screening was 500 p.p.m. With the combined technique, it was possible to directly detect a target gene at 100 p.p.m.

DISCUSSION

The idea of 'negative amplification', in which tester sequences are converted to counterpart driver sequences, was new when published in 1999. But the DSC approach turned out to be not reproducible by us. We tried and failed to employ this system and we hypothesized that the DSC might have weakness due to the construction of the adapters. We designed new adapters/primers to make the subtraction system more robust and called the resulting method NSC, with reference to its DSC-inspired development.

With SSH and NSC, differential gene expression can be analysed in a fast, robust and sensitive manner. It takes less than two weeks to apply this combined technique to successfully identify differentially expressed genes in *Solanum*. Because of its sensitivity, the combined technique is a significant improvement in comparison to SSH, the only commercially available method for transcriptome subtraction. With *Solanum* as a model organism, we have shown that the combined technique is capable of directly identifying differentially expressed genes of two different plants. With SSH the problem of false positive occurs, in which subtraction products of highly complex eukaryotes contain many

background sequences which mask the target. The sensitivity of the combined technique to detect differentially expressed genes, 100 p.p.m, is just the expected concentration of mRNA of the middle abundance class in highly complex eukaryotes (8). With SSH, it is only possible to directly detect differentially expressed genes of the highest abundance class, whereas with the combined technique sequences in the middle abundance class, to which most of the differentially expressed genes are expected to belong (9), can also be identified directly. This sensitivity is comparable with that of another SSH helper called mirror orientation selection (MOS) (10). In addition, MOS seems to be slightly easier to perform because only one round of subtraction is applied. On the other hand, NSC produces less false positive clones, because the progressive subtraction was applied for more rounds. Although our experiments were carried out on *Solanum*, this combined technique is expected to be suitable for the analysis of the differential gene expression in other highly complex eukaryotes as mammalians, whose genome complexity is comparable or lower than that of potatoes. In principle, this combined method can identify any target. All adapters/primers applied (also during cDNA synthesis) are equipped with the recognition sequence of the same restriction endonuclease, therefore every complete cDNA molecule contains at least two of these recognition sites, which makes every cDNA prone to digestion with the restriction enzyme.

Recently, microarray techniques came up and are now widely used to detect differential gene expression. Not to deny their many advantages, these techniques are expensive, for many organisms not yet standardized. The main advantage of the combined technique with SSH/NSC is its ability to isolate differentially expressed genes quickly and cost-efficiently from non-standard models, for those microarrays are unavailable.

ACKNOWLEDGEMENTS

We thank Dr Walter Taylor for correcting the English. Funding to pay the Open Access publication charges for this article was provided by Deutsche Forschungsgemeinschaft.

Conflict of interest statement. None declared.

REFERENCES

1. Liang,P., Zhu,W., Zhang,X., Guo,Z., O'Connell,R.P., Averboukh,L., Wang,F. and Pardee,A.B. (1994) Differential display using one-base anchored oligo-dT primers. *Nucleic Acids Res*., **22**, 5763–5764.
2. Heller,M.J. (2002) DNA microarray technology: devices, systems, and applications. *Annu. Rev. Biomed. Eng*., **4**, 129–153.
3. Velculescu,V.E., Zhang,L., Vogelstein,B. and Kinzler,K.W. (1995) Serial analysis of gene expression. *Science*, **270**, 484–487.
4. Lisitsyn,N. and Wigler,M. (1993) Cloning the differences between two complex genomes. *Science*, **259**, 946–951.
5. Diatchenko,L., Lau,Y.F., Campbell,A.P., Chenchik,A., Moqadam,F., Huang,B., Lukyanov,S., Lukyanov,K., Gurskaya,N., Sverdlov,E.D. and Siebert,P.D. (1996) Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proc. Natl Acad. Sci. USA*, **93**, 6025–6030.
6. Luo,J.H., Puc,J.A., Slosberg,E.D., Yao,Y., Bruce,J.N., Wright,T.C.,Jr, Becich,M.J. and Parsons,R. (1999) Differential subtraction chain, a method for identifying differences in genomic DNA and mRNA. *Nucleic Acids Res*., **27**, e24.
7. Chomczynski,P. and Sacchi,N. (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem*., **162**, 156–159.
8. Bishop,J.O., Morton,J.G., Rosbash,M. and Richardson,M. (1974) Three abundance classes in HeLa cell messenger RNA. *Nature*, **250**, 199–204.
9. Hastie,N.D. and Bishop,J.O. (1976) The expression of three abundance classes of messenger RNA in mouse tissues. *Cell*, **9**, 761–774.
10. Rebrikov,D.V., Britanova,O.V., Gurskaya,N.G., Lukyanov,K.A., Tarabykin,V.S. and Lukyanov,S.A. (2000) Mirror orientation selection (MOS): a method for eliminating false positive clones from libraries generated by suppression subtractive hybridiztion. *Nucleic Acids Res*., **28**, e90.