Identification of fragments of human transcripts from a defined chromosomal region: representational difference analysis of somatic cell hybrids

Peter C. Groot and Bernard A. van Oost*

Department of Clinical Sciences of Companion Animals, Faculty of Veterinary Medicine, University of Utrecht, PO Box 80154, 3508 TD Utrecht, The Netherlands

Received May 21, 1998; Revised August 12, 1998; Accepted August 18, 1998

ABSTRACT

We have tested representational difference analysis of cDNAs from somatic cell hybrids as a means to directly isolate expressed sequences derived from a defined chromosomal region. To this end, the hamster–human somatic cell hybrid Q1Z, which carries Xq28 as the only human chromosomal fragment, was used as Tester and the parental hamster cell line Y21 as Driver. After two rounds of subtraction, two major products of 510 and 307 bp were obtained, derived from the highly expressed human Xq28-derived QM gene and from a hamster repeat sequence strongly up-regulated in Q1Z, respectively. In a second subtraction experiment these fragments were added to the driver, to prevent their reappearance. After three rounds of subtraction a more complex difference product was obtained. Of 26 different fragments analysed, 12 fragments were derived from Xq28-derived genes, 10 of which were from known genes. One fragment was derived from a hamster gene strongly up-regulated in Q1Z. These results demonstrate that cDNA RDA can be used to isolate gene fragments from defined chromosomal regions and that suppression with major products, derived from highly expressed genes, is advantageous to isolate larger number of fragments, presumably derived from rarer transcripts.

INTRODUCTION

Representational difference analysis of cDNAs (cDNA RDA) (1) is an effective method to identify fragments of differentially expressed genes (2–7). This technique has been shown to be highly sensitive and was used to identify genes which are expressed in only a very small fraction of the cells from which the Tester was derived (1). This implies that genes with very low levels of expression should also be amenable to isolation, provided they are not expressed in the cells used to prepare the Driver. Therefore, using cDNA RDA, it should be possible to isolate genes from defined chromosomal fragments present in somatic cell hybrids, by comparing these hybrids with the parental cell lines from which they are derived.

To test this possibility, we have applied cDNA RDA to the somatic cell hybrid Q1Z (8), by comparing it with the Chinese hamster ovary cell line Y21 (9), from which it was derived. Cell hybrid Q1Z carries the region of the X chromosome distal to the Fra(X) site at Xq27.3 as the only human chromosomal fragment. The Q1Z cell line was previously used for the isolation of Xq28-derived genes by screening cDNA libraries with Q1Z-derived cosmid clones. This has resulted in the isolation of QM, which is highly expressed in Q1Z (10), and RENBP (11). The Xq27–28 region is one of the most gene dense regions of the human genome (12) and harbour a large number of genes implicated in human disease (13).

In this report we demonstrate that cDNA RDA can be used effectively to isolate gene fragments from defined chromosomal regions. Suppression with major products from highly expressed genes appeared to be advantageous for the isolation of larger number of fragments from less abundant transcripts.

MATERIALS AND METHODS

Oligonucleotides

Sequences of oligonucleotides used in cDNA RDA were as follows: R-Bgl-12, 5′-AGCACCTTCAGCCTCTCACCAGCA-3′; R-Bgl-12, 5′-ATGTGGCGGTAG-3′; J-Bgl-12, 5′-GATCTGGATTGTTCCA-3′; J-Bgl-24, 5′-GGCACGTCGACTATCACTCGATCGAA-3′; J-Bgl-24, 5′-GATCTGTTCTATGCTTGA-3′; N-Bgl-24, 5′-GATCTGTTCTATGCTTGA-3′; N-Bgl-12, 5′-AGGCCATGTTGCATCCGAGGGAA-3′; N-Bgl-12, 5′-GATCTCCCTCG-3′ (1).

Cell lines, mRNA and cDNA isolation

The hamster–human somatic cell hybrid Q1Z (8) and the parental hamster cell line Y21 (9) were kindly provided by Dr S. Warren. Cells were grown in Dulbecco’s modified Eagle’s medium (DMEM), 10% fetal calf serum, 50 µg/ml gentamycin in a humidified atmosphere of 5% CO2 at 37°C. Poly(A)+ RNA was isolated (Fast-Track 2.0 kit, Invitrogen) and converted into double-stranded cDNA (Copy kit, Invitrogen). In order to avoid differences between samples caused by variations in growth conditions or downstream processing, Q1Z and Y21 were grown side by side and mRNA isolation and cDNA construction were done in parallel. The quality of the cDNAs obtained was tested by
hybridization with a probe for the QM (10) and β-actin genes, respectively.

**cDNA RDA**

Except for a few modifications, the protocol of Hubank and Schatz (1), which was based on the original protocol of Lisitsyn et al. (14), was followed. Detailed protocols were kindly provided by N. Lisitsyn and D. Schatz, respectively. Briefly, cDNA was digested with the four-cutter DpnII (GATC) to obtain maximum representation, ligated to R-BglII adaptors and amplified by PCR to generate amplicons. Adaptors were removed by digestion with DpnII followed by S300-HR spin column chromatography (Pharmacia). To generate the Tester, J-BglII adaptors were ligated to the Q1Z amplicons and excess adaptors were removed using S300 HR spin columns. After subtractive hybridization, the hybridization product was re-amplified as described, to generate DP1, the first difference product. After changing J-BglII adaptors for N-BglII adaptors, a new round of subtractive hybridization was performed to generate DP2, etc. In the first round of subtractive hybridization, a Driver:Tester ratio of 100:1 was used, in the second round 800:1 and in the third round 40 000:1.

PCR was always performed for 20 cycles on a PCT-100 Thermal Cycler (MJ Research). Toq DNA polymerase was obtained from Gibco BRL. Prior to the preparative PCR reactions, analytical PCR reactions were performed to determine the proper amount of input DNA (Fig. 1 and Discussion). To this end, 30 µl PCR reactions of serial dilutions of the input DNA were carried out and of each PCR, 20 µl was analysed on 2% agarose.

Normalization of difference products was achieved as follows. Plasmid DNA was isolated using Wizard minipreps (Promega) and sequenced using M13 sequencing primers, according to the instructions of the manufacturer (Pharmacia). Southern blots with Tester and Driver amplicons were analysed by PCR with M13 sequencing primers, digested with DpnII and analysed on gel. Based on the yields obtained, amounts of each PCR were pooled in such a way that a ‘normalized’ difference product was obtained, which was then subcloned.

**Suppression with difference products**

In order to suppress the appearance of difference products isolated in previous subtraction experiments, these fragments were amplified from the pZero clones with M13 primers, digested with DpnII and 500 ng each were added to the Driver.

**RESULTS**

In order to isolate human Xq28-derived gene fragments, cDNA RDA was performed according to the procedure of Hubank and Schatz (1). To determine the efficacy of the procedure, hybridization with human QM cDNA was used as a specific assay in the first cDNA RDA experiment. The QM gene is specifically expressed in the somatic cell hybrid Q1Z (10). To determine the efficacy of the procedure, hybridization with QM cDNA was used as a specific assay in the first cDNA RDA experiment. The QM gene is specifically expressed in the somatic cell hybrid Q1Z (10).
The first important step of the RDA is the generation of amplicons by PCR of linker-ligated DpnII-digested cDNA fragments. Titration to determine the proper amount of input DNA for the preparative PCR reactions was achieved by performing a series of so-called analytical PCR reactions in which the amount of input DNA was varied (16). Seven PCR reactions with 1:5 serial dilutions of linker-ligated cDNA fragments as input enabled us to scan a concentration range spanning four orders of magnitude (from ∼1000 to 0.064 pg/µl; Fig. 1). Ethidium bromide staining (Fig. 1A) of the titration of Driver (lanes 1–7) and Tester (lanes 8–14) showed that with increasing amounts of input DNA the amount of amplicons produced increased initially (lanes 3–7 and 9–14) and then levelled off (lanes 2–3 and 8–9) or even seemed to decline (compare lanes 1 and 2). Upon hybridization with the QM probe (Fig. 1B) the expected 550 bp QM band (510 bp + linkers) appeared in Tester lanes 8–14. Very weak hybridization with a band of similar length in the Driver could also be detected. This is likely the result of very low expression of the hamster QM gene in Y21, which was also detected upon hybridization of the QM probe to the Y21 cDNA (not shown). Using northern analysis this expression had not previously been detected (10). Remarkably, in lanes 8 and 9, a second, strongly hybridizing band of ∼350 bp appeared, which in lane 8 is almost as strong as the cognate 550 bp band. The smaller fragment disappears after re-amplification (not shown) and therefore must represent a different conformation of the 550 bp band, most likely a single-stranded form which has folded back on itself due to the perfect inverted repeats present on both ends of each molecule, forming so-called panhandles, a phenomenon described before (17). Close inspection of hybridization with the actin probe (Fig. 1) showed a similar effect: in lanes 1–3 and 8–9 the appearance of a hybridizing band of ∼260 bp can be seen. A fragment of this length is not present in the other lanes. Together, this indicated exhaustion of the PCR and this would also explain the apparent reduction in the amount of DNA produced, as shown in lane 1, resulting from less efficient binding of ethidium bromide to a larger relative amount of single-stranded DNA. The presence of single-stranded DNA can also explain why complete removal of adaptors is apparently not possible and why the change of adaptors between successive rounds of subtractive hybridization is absolutely necessary for RDA to actually work (for without the change of adaptors this single-stranded DNA, consisting mainly of unwanted background, would survive subtractive hybridization). Based on the experiment shown in Figure 1, the optimal amounts of input DNA used for the PCR would be the amounts used in lanes 4 and 10 for Driver and Tester, respectively.

The result of the first subtraction experiment is shown in lanes 1–4 of Figure 2. Lanes 1 and 2 show Driver and Tester and lanes 3 and 4 the first and second difference products (DP1 and DP2), respectively. DP1 shows few discrete bands with a background smear ranging from 150 to 1000 bp. DP2 shows two major products, MP1 of 560 bp and MP2 of 350 bp, against a faint background smear. Digestion with MspI revealed that MP1 represented the QM band, as this band was completely cut into fragments of the lengths predicted by the sequence of the QM gene (not shown). Hybridization analysis with probes for QM and actin confirmed the effect of the subtraction: QM was strongly enriched and actin was depleted upon subtraction (not shown). As we detected only two major products we decided not to perform a third round of subtraction. Instead, we subcloned DP2 in order to reveal the identity of MP2.

Contrary to expectations, DP2 appeared to be complex. One hundred and seven randomly selected clones were analysed by T-tracking and ∼50 clones were sequenced. From the 123 DpnII fragments present in these clones (16 clones contained multiple DpnII fragments as a result of co-ligation during subcloning), 17 fragments were identical to MP1 and 27 fragments were more or less identical to MP2 (see below). The remaining 79 fragments were unique, except for three fragments which were found twice and one fragment which was found three times.
The MP2 fragment hybridized strongly to Q1Z amplicons, but no detectable hybridization to Y21 amplicons could be observed (Fig. 3A). However, hybridization to BamHI-digested Y21 and Q1Z genomic DNA resulted in identical, repeat-like hybridization patterns (Fig. 3B), indicating that MP2 was hamster derived. T-tracking revealed multiple differences between any two MP2 fragments compared, indicating that MP2 must be derived from different copies of a nearly identical hamster sequence, which is strongly up-regulated in Q1Z. Using BLASTN, no homology of MP2 with any known gene could be detected.

Four additional Xq28-derived cDNA fragments were identified on the basis of sequence homology with published sequences. Two fragments of 200 and 440 bp were derived from the actin binding protein gene Filamin (FLN1), one 350 bp fragment from the GDP dissociation inhibitor gene (XAP-4) and one 250 bp fragment from the iduronate 2-sulfatase gene (IDS). That these fragments were indeed derived from human Xq28 was experimentally confirmed for both FLN1 fragments by hybridization to BamHI-digested genomic DNA of Y21 and Q1Z. In both cases hybridization was observed to a 10 kb band present in Q1Z only (not shown).

For all other cDNA fragments, BLAST searches either revealed some degree of homology with various known genes or did not detect any homology with known sequences. In order to reveal whether fragments were of human or hamster origin, 43 different fragments were hybridized against Driver and Tester amplicon blots. Nine fragments failed to hybridize. Twenty-three fragments hybridized to both Tester and Driver and were thus derived from the hamster gene ‘background’. Three of these cDNA fragments hybridized much more strongly to Tester than to Driver and most likely represent hamster genes which were up-regulated in Q1Z. For 11 fragments, significant hybridization could be detected to Tester amplicons only. However, hybridization of five of these fragments to genomic Southern blots showed that they all hybridized to the same bands in both Y21 and Q1Z and thus also represented up-regulated hamster-derived genes.

In order to increase the yield of Xq28-derived sequences, a second subtraction experiment was performed in which the two main products of the first subtraction experiment were added to the driver in order to prevent their reappearance. Also, three rounds of subtraction were performed. In Figure 2 the difference products obtained after each round of subtraction are shown in lanes 5–7, respectively. Both DP2 and DP3 (lanes 6 and 7) were more complex than DP2 of the first subtraction experiment (lane 4). DP3 consisted of a number of discrete bands differing in intensity. For further analysis DP3 was ‘normalized by PCR’ (see Materials and Methods for details). After normalization, ~15 discrete bands were visible after ethidium bromide staining. The normalized product was subcloned and 26 clones, with inserts ranging from 65 to 558 bp, were sequenced. Ten clones contained Xq28-derived gene fragments as they showed complete identity to fragments of human cDNAs known to be derived from Xq28: four fragments (558, 526, 279 and 75 bp long) corresponded to FLN1, one fragment (314 bp) to XAP-4, one fragment (348 bp) to IDS, two fragments (276 and 111 bp) to the L1-CAM gene and two fragments (524 and 393 bp) to the NAD(H) isocitrate dehydrogenase γ-subunit precursor gene (IDH) (Fig. 4). Two fragments (215 and 440 bp) yielded no hit in BLAST searches but were identified as Xq28-derived cDNA fragments as they hybridized to Tester but not to Driver on amplicon blots as well as to Q1Z but not to Y21 DNA on genomic blots. One other fragment hybridized strongly with Tester amplicons only, but hybridized to identical bands in genomic Y21 and Q1Z DNA, demonstrating that it represented a strongly up-regulated hamster gene. Four fragments hybridized to Tester as well as to Driver amplicons and thus represented hamster-derived cDNA fragments, one of which appeared to be up-regulated. The remaining eight fragments failed to hybridize and because they yielded no significant hits in BLAST searches of available databases, these fragments were not investigated further. A summary of the results of the two subtraction experiments performed is given in Table 1.
Table 1. Analysis of difference products

<table>
<thead>
<tr>
<th>Origin (analysis of 123 randomly chosen DP2 fragments)</th>
<th>n</th>
<th>Clone</th>
<th>Size (bp)</th>
<th>Derived from</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human Xq28</td>
<td>21</td>
<td>MP1</td>
<td>510</td>
<td>QM</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DP2-54</td>
<td>396</td>
<td>Filamin</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DP2-45</td>
<td>162</td>
<td>Filamin</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DP2-52</td>
<td>317</td>
<td>XAP-4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DP2-75</td>
<td>204</td>
<td>IDS</td>
</tr>
<tr>
<td>Hamster up-regulated</td>
<td>40</td>
<td>MP2</td>
<td>308</td>
<td>? (accession no. AJ009564)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DP2-6</td>
<td>290</td>
<td>? (accession no. AJ009566)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DP2-39</td>
<td>215</td>
<td>? (accession no. AJ009565)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DP2-8A</td>
<td>208</td>
<td>? (accession no. AJ009567)</td>
</tr>
<tr>
<td>Hamster background</td>
<td>19</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Origin unknown</td>
<td>43</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Experiment 2 (analysis of 26 different DP3 fragments)

<table>
<thead>
<tr>
<th>Origin (analysis of 26 different DP3 fragments)</th>
<th>n</th>
<th>Clone</th>
<th>Size (bp)</th>
<th>Derived from</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human Xq28</td>
<td>12</td>
<td>DP3-1</td>
<td>562</td>
<td>Filamin</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DP3-2</td>
<td>524</td>
<td>Filamin</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DP3-13</td>
<td>282</td>
<td>Filamin</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DP3-27</td>
<td>62</td>
<td>Filamin</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DP3-5</td>
<td>524</td>
<td>IDH</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DP3-9</td>
<td>393</td>
<td>IDH</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DP3-10</td>
<td>348</td>
<td>IDS</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DP3-14</td>
<td>276</td>
<td>L1-CAM</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DP3-25</td>
<td>111</td>
<td>L1-CAM</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DP3-11</td>
<td>317</td>
<td>XAP-4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DP3-6</td>
<td>440</td>
<td>? (accession no. AJ009562)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DP3-16</td>
<td>220</td>
<td>? (accession no. AJ009563)</td>
</tr>
<tr>
<td>Hamster up-regulated</td>
<td>2</td>
<td>DP3-12</td>
<td>300</td>
<td>? (accession no. AJ009568)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DP3-15</td>
<td>250</td>
<td>? (accession no. AJ009569)</td>
</tr>
<tr>
<td>Hamster background</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Origin unknown</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In experiment 1, DP2 was subcloned and 107 random clones, comprising 123 DpnII fragments were analysed. In experiment 2, DP3 was normalized by PCR before subcloning and 26 different DpnII fragments were analysed (see Materials and Methods). The origin of the cloned DpnII fragments was determined by T-tracing to identify identical clones (only in experiment 1), by hybridization of labelled fragments to Southern blots with Tester and Driver amplicons or with BamHI-digested Y21 and Q1Z genomic DNA and/or by sequencing. The sequences of the fragments derived from known Xq28-derived genes were identical to fragments of the published sequences with accession nos M64241 (QM), X53416 (FLN1), X79353 (XAP-4), M58342 (IDS), Z68907 (IDH) and X58776 (L1-CAM). In the second column (n), the number of fragments found in each class is indicated. For MP1 and MP2, n indicates how many times these fragments were found. Only clones harbouring human Xq28 and hamster up-regulated gene fragments which had been sequenced are given in the table. Accession numbers for the sequences which have been deposited in the EMBL database are indicated.

DISCUSSION

Using cDNA RDA (1) we have been able to isolate fragments of genes from a defined chromosomal region. Our results indicate that cDNA RDA can be very useful for this purpose, especially because it does not require genomic DNA clones covering the chromosomal region of interest, as is necessary for the direct selection of cDNAs (18,19). This was also shown by a concurrent study of Tajima et al. (20), who were able to identify chromosome 17-specific cDNAs using human–mouse somatic hybrid cell lines.

Once cDNA RDA is operational, performing it is straightforward and a difference product can be obtained in <2 weeks. In our experience, the most time-consuming part of cDNA RDA is the analysis of the difference products obtained. But this is equally true and even more so for other techniques used for the identification of differentially expressed sequences, such as differential display PCR (21) or AFLP-based mRNA fingerprinting (22). In comparison, cDNA RDA allows for much easier cloning of difference products and does not require testing of many different primer combinations and many gel runs. Moreover, these techniques suffer from a bias against rare mRNAs (23,24). In principle, cDNA RDA is also sensitive to this bias, but iteration greatly increases its sensitivity (see below).

Currently, comprehensive techniques for the analysis of differential gene expression have been developed, such as serial analysis of gene expression (25) and hybridization to microarrays (26) or to high density oligonucleotide arrays (27). The latter two techniques require prior sequence information and all three techniques require special equipment and are (still) very expensive. In comparison, cDNA RDA is relatively inexpensive and can also be carried out by small laboratories.
Few small, but in our view important, changes to the protocols developed by Lisitsyn et al. (14) and Hubank and Schatz (1) were made. We incorporated so-called analytical PCR reactions prior to each preparative PCR reaction, which greatly increased the speed and predictability of the procedure. For removal of adaptors we used spin column chromatography instead of purification of amplicons from an agarose gel, which is a potential source of contamination. This procedure was fast and worked well in our hands.

Our first experiment yielded only two major products, the Xq28-derived QM fragment and the MP2 fragment derived from a hamster gene or gene family which was strongly up-regulated in Q12.

Repeating the experiment after addition of these two products to the Driver completely suppressed the reappearance of MP1 and MP2 and yielded 14 additional subtraction products, 12 of which came from Xq28 (Table 1), clearly demonstrating the effectiveness as well as the necessity of iteration. To appreciate this, the effect of kinetic enrichment must be considered. Kinetic enrichment is likely the most important factor in achieving the very high enrichment factors (up to 1 000 000-fold after three cycles of subtractive hybridization) observed in RDA (14). In each round of subtractive hybridization, the kinetic enrichment of a given fragment is proportional to the square of its concentration. Initial small differences in concentration will thus become very large after a few rounds of RDA. This has profound implications in the case of cDNA RDA, where amplicons are generated from mRNA populations consisting of mRNA species present at greatly differing concentrations. Because of these differences, cDNA RDA is expected to result in the recovery of only those differentially expressed fragments which were initially present at the highest concentration. Other fragments are lost, despite the fact that the latter fragments may represent genuine subtraction products. The isolation of only two major products, derived from highly expressed genes, in our first experiment can be explained in this way. Without iteration, this would be considered a very disappointing result. However, the isolation of 14 additional subtraction products against a low background in the second experiment clearly demonstrates the effectiveness of this approach. This is in agreement with previously published results (1,5,7).

From four Xq28 genes, IDH, IDS, L1-CAM and FLN1, more different DpnII fragments were isolated (Fig. 4). However, many more fragments could have been isolated from these genes. Although some of these fragments may have been present in the difference products obtained, we consider it likely that most of them did not survive up to DP3 for stochastic reasons.

In our view, the most effective future improvement of cDNA RDA would be the simultaneous construction of a cDNA library from the cDNA population used to generate the Tester. After each cycle of cDNA RDA, this library should then be screened with the normalized DP3 or perhaps even DP4 (DP2 will likely still contain too much background), to yield full-length cDNAs rather than cDNA fragments. This would make identification of the genes involved much easier and would also make suppression with previously identified products much more effective. For instance, addition of the two FLN1 fragments from our first experiment to the Driver would not have suppressed the appearance of any of the four FLN1 fragments in the second experiment, whereas addition of the complete FLN1 cDNA would have suppressed them all.

ACKNOWLEDGEMENTS

We gratefully acknowledge Dr N. Lisitsyn and Dr D. Schatz for providing detailed protocols and Dr M. Schutte for advice. This work was supported by the Dutch Foundation for Scientific Research (NWO) (project-grant no. 901-04-136).

REFERENCES