Expression profiling of single cells using 3' prime end amplification (TPEA) PCR

Alistair K. Dixon¹,³, Peter J. Richardson¹,², Kevin Lee³, Nigel P. Carter¹ and Tom C. Freeman¹, *

¹The Sanger Centre, The Wellcome Genome Campus, Hinxton, Cambridge CB10 1SA, UK, ²Department of Pharmacology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QJ, UK and ³Parke-Davis Neuroscience Research Centre, Cambridge University Forvie Site, Robinson Way, Cambridge CB2 2QB, UK

Received June 11, 1998; Revised and Accepted August 21, 1998

ABSTRACT

The ability to relate the physiological status of individual cells to the complement of genes they express is limited by current methodological approaches for performing these analyses. We report here the development of a robust and reproducible method for amplifying 3' sequences of mRNA derived from single cells and demonstrate that the amplified cDNA, derived from individual human lymphoblastoma cells, can be used for the expression profiling of up to 40 different genes per cell. In addition, we show that 3' prime end amplification (TPEA) PCR can be used to enable the detection of both high and low abundance mRNA species in samples harvested from live neurons in rat brain slices. This procedure will facilitate the study of complex tissue function at the cellular level.

INTRODUCTION

DNA sequence generated by genome and expressed sequence tag (EST) sequencing projects will provide the basis to further our understanding of the control and mode of action of individual gene products. In this respect, analysis and comparison of when, where and to what degree genes are expressed, commonly known as expression profiling, is certain to play an essential role in the functional characterization of newly identified genes (1,2).

Many tissues, such as the immune and nervous systems, are composed of highly heterogeneous cell populations. A key factor in understanding their physiology, and the role of specific gene products expressed within them, will be the ability to examine gene usage in the context of this cellular diversity. In the past, methods such as northern blotting and nuclease protection assays were employed to study gene expression. More recently, a number of powerful new technologies have been developed for assessing simultaneously the expression of large numbers of genes (3,4). All these techniques, however, require relatively large amounts of RNA and currently lack the sensitivity to analyse specimens derived from small populations of cells or indeed from an individual cell.

At present, methodologies for the analysis of gene expression within single cells are limiting. Whilst in situ hybridization provides detailed information on a gene's cellular expression pattern in intact tissue, be it whole-mounts or tissue sections, the technique is relatively laborious to perform and is incapable of analysing more than a very small number of transcripts in a single preparation. In recent years, the polymerase chain reaction (PCR) has been used successfully to investigate gene expression in cytoplasmic samples derived from single cells (5). The nested-primer approach has good sensitivity, but the analysis is restricted to a small number of closely related genes from specific gene families (6). Other techniques capable of detecting the expression of unrelated genes in a single cell include T7 RNA polymerase amplification of mRNA (7,8), and PCR after prior homopolymeric tailing of the first strand cDNA (9–11). However, neither of these approaches have been demonstrated to be capable of analysing more than a small number of genes and are also not widely used. The former is technically difficult, whilst the latter may be biased against long transcripts and often requires subsequent cloning of the amplified products. There exists therefore a great need for improved methodologies that can quickly, easily and sensitively analyse gene expression at the cellular level.

Here we report the development of an approach, called 3' prime end amplification (TPEA) PCR, for the analysis of gene expression in small samples or single cells, which overcomes many of the technical limitations of existing methodologies. In this method, the 3' region of mRNA is amplified arbitrarily by PCR using a novel combination of primers. The amplified cDNA, which represents the most diverse region of gene sequence, can then be analysed by a second round of PCR using gene-specific primers. We demonstrate that it is possible to analyse the expression of up to 40 genes (20 in duplicate) in single human lymphoblastoma cells. We also show that this method is ideally suited to the analysis of genes expressed at low levels in small populations of cells, by demonstrating the expression of the adenosine A₂a receptor in cholinergic neurons of the rat striatum.

*To whom correspondence should be addressed. Tel: +44 1223 494709; Fax: +44 1223 494919; Email: tcf@sanger.ac.uk
MATERIALS AND METHODS

Lymphoblastoma cell sorting

An Epstein-Barr virus transformed lymphoblastoid cell line (HRC575, ECACC, Porton Down, UK) was maintained in log phase growth in RPMI 1640 medium supplemented with 16% foetal calf serum, 2 mM l-glutamine and penicillin–streptomycin (100 U/ml and 100 µg/ml, respectively). Cells at ∼10⁶/ml were stained with the bisbenzimadazole dye Hoechst 33342 (Sigma, Poole, UK) at 1 µg/ml for 30 min at 37°C. Cells were sorted by using the Autoclone attachment of a Coulter Elite ESP flow cytometer, 300 mW of all lines UV from a Coherent 306 laser and by using single drop and complete abort sorting settings. Time of flight, forward and right angle scatter and Hoechst fluorescence peak and area measurements were used to ensure the sorting of single cells. The accuracy of sorting (both spatial and numerical) was tested by sorting single fluorescent beads (DNA Check, Coulter Corp.) into 96-well plates and viewing the plates on a fluorescence microscope.

Reverse transcription and cDNA amplification

Lymphoblastoma cells were sorted into 96-well plates containing 7 µl of freshly prepared lysis buffer [50 mM Tris–HCl (pH 8.3), 75 mM KCl, 3 mM MgCl₂, 5 mM NP-40 (Sigma) and 1.5 U of RNase inhibitor (Pharmacia, Milton Keynes, UK)]. This buffer leaves the nucleus intact (11). After 5 min on ice, nuclei were removed by centrifugation (8000 g, 5 min at 4°C), the supernatant aspirated, and the RNA reverse transcribed in a reaction volume of 10 µl containing: 1 x first-strand buffer, 200 U M-MLV reverse transcriptase (Gibco BRL, Paisley, UK), 0.5 ng reverse transcription primer for 60 min at 37°C. The RT primer was composed of an anchored oligo(dT) primer with a specific 5' heel sequence: CTC TCA AGG ATC TTA CCG CTT TTT TTT TTT TTT (A,G,C).

Second-strand cDNA synthesis was initiated by incubation of the first-strand cDNA with 1 ng of a primer consisting of (5' to 3') a 20 base sequence selected due to its absence from the mammalian databases, a stretch of five random nucleotides and an anchored oligo(dT) primer with a specific 5' heel sequence: CTC TCA AGG ATC TTA CCG C (Sigma). An arbitrary primer closest to the 5' heel sequence of the oligo(dT) primer will undoubtedly prime second-strand DNA synthesis at many sites on the first strand cDNA, the subsequent PCR amplification, and 2.5% of the product used in each gene specific PCR reaction.

Extracting neuronal contents

Coronal slices (300 µm) from 14–28 day-old male Sprague–Dawley rats containing the striatum were viewed with a Zeiss Axioskop microscope (Carl Zeiss Ltd, Welwyn Garden City, UK) fitted with a × 64 water-immersion objective lens together with gradient contrast optics (Luigs and Neumann, Ratingen, Germany). Light in the infrared range (∼740 nm) was used in conjunction with a contrast-enhancing Newvicom camera (Hamamatsu, Hamamatsu City, Japan) to resolve individual neurons within slices (27). The cytoplasm from large cells (>30 µm in one dimension) was gently aspirated under visual control into a patch-clamp recording electrode until at least 40% of the somatic cytoplasm had been collected. The electrode was then withdrawn from the cell to form an outside-out patch which prevented contamination when the electrode was subsequently withdrawn from the slice. The contents of the electrode were forced into a microtube and reverse transcribed, subjected to 3' cDNA amplification, and 2.5% of the product used in each gene specific PCR reaction.

RESULTS

TPEA-PCR

Assay development was performed on lymphoblastoma cells in the G₂/M phase of the cell cycle. Groups of 100, 10 and single cells were flow sorted into wells containing lysis buffer and the mRNA reverse transcribed. A proportion of the sorted cells then underwent 3’ end amplification, as described in Materials and Methods (Fig. 1). Gene-specific PCR assays for eight ‘housekeeping’ genes were carried out on lymphoblastoma cDNA, before and after cDNA amplification. Following reverse transcription only, the expression of each of the housekeeping genes could be detected when cDNA generated from between one and 100 cells was used in each PCR assay (Fig. 2). It was not possible, however, to detect the expression of any of these genes when cDNA equivalent to less than one cell was assayed. Following TPEA-PCR, however, the expression of all eight of the genes was detectable even when amplified cDNA generated from the equivalent of as little as 5% of one cell was present in the PCR assay (Fig. 2).
Table 1. Primer sequences used for gene-specific assays

<table>
<thead>
<tr>
<th>Gene</th>
<th>Symbol</th>
<th>Acc. No.</th>
<th>Forward</th>
<th>Reverse</th>
</tr>
</thead>
<tbody>
<tr>
<td>Riboproopin L5</td>
<td>RPL5</td>
<td>U14966</td>
<td>GTAACCGAATCAACGAGCATG</td>
<td>CCGCTCTTGAGATCGTCAG</td>
</tr>
<tr>
<td>Riboproopin L2</td>
<td>RPL2</td>
<td>U14967</td>
<td>GTGCTTGTGACGCAATGATGC</td>
<td>TTTCAGGTTGCAGGTCAG</td>
</tr>
<tr>
<td>Riboproopin L27a</td>
<td>RPL27a</td>
<td>U14968</td>
<td>GAAAGTTGAGGATGCTGGCA</td>
<td>AGTGGTCATGACCCAG</td>
</tr>
<tr>
<td>Riboproopin L28</td>
<td>RPL28</td>
<td>U14969</td>
<td>ATCAACAGATGTCGCCG</td>
<td>TGACATTGTTCGGGTGGGG</td>
</tr>
<tr>
<td>Riboproopin S5</td>
<td>RPS5</td>
<td>U14970</td>
<td>GTGAAACGAGGATGCTGGCA</td>
<td>CCAAGGCGAGACAGGTTA</td>
</tr>
<tr>
<td>Riboproopin S9</td>
<td>RPS9</td>
<td>U14971</td>
<td>CTATTCGCGGGATGCTGCC</td>
<td>TGTTTATTTCGAGGAAAGAG</td>
</tr>
<tr>
<td>Riboproopin S10</td>
<td>RPS10</td>
<td>U14972</td>
<td>GGAGATTGCTGGCCACAC</td>
<td>AAAGTGATAGCCTGGGTC</td>
</tr>
<tr>
<td>Riboproopin S29</td>
<td>RPS29</td>
<td>U14973</td>
<td>ATGTGGCAAGGCTTCCTC</td>
<td>TTTTTGAGTATGATGCCC</td>
</tr>
<tr>
<td>Housekeeping protein G257</td>
<td>HS2PQ7</td>
<td>M18106</td>
<td>GAAATCAGGCTGGAGAAGC</td>
<td>GGAGTGCGAAGTACGTC</td>
</tr>
<tr>
<td>Glutamyl-hydroxamate-3-</td>
<td>G-3-PDH</td>
<td>M31169</td>
<td>CGCAAGCTTGGCAAGAAGC</td>
<td>AGGAGTCGCAAGATGAC</td>
</tr>
<tr>
<td>Cytoplasmic beta-actin</td>
<td>ACTB</td>
<td>A800467</td>
<td>CTGTGACATGTCATGAAAGG</td>
<td>ACATGTCGGAAGATGAC</td>
</tr>
<tr>
<td>JUND</td>
<td>JUND</td>
<td>X56681</td>
<td>CTCGCAAGAACAAAGCTTT</td>
<td>ACAACAGAAGATGAC</td>
</tr>
<tr>
<td>Elongation factor 1</td>
<td>EF1</td>
<td>M18288</td>
<td>CCAAGAAGACCAATGAGGCC</td>
<td>TGCTGAGACATTTTGTG</td>
</tr>
<tr>
<td>Cell division cycle 25B</td>
<td>CDC25B</td>
<td>M18104</td>
<td>AGTCGATCGATCAGAACA</td>
<td>CCCCACTCTGAGATACAG</td>
</tr>
<tr>
<td>CDC19</td>
<td>CDC19</td>
<td>X13512</td>
<td>CAAACGCTGAGGACATGTT</td>
<td>GGAATACAAAGAGGACTG</td>
</tr>
<tr>
<td>CDC79A</td>
<td>CDC79A</td>
<td>L32754</td>
<td>CTTCGCGGGCTTCCTGAG</td>
<td>TTAGAGATTGCGGGAGATTG</td>
</tr>
<tr>
<td>Immunoglobulin M</td>
<td>IM</td>
<td>X17115</td>
<td>GTCCCGGATGAGAATGTCG</td>
<td>GAGCTGTCGACGGTCTC</td>
</tr>
<tr>
<td>CD2</td>
<td>CD2</td>
<td>M16445</td>
<td>TTCTGAACTCTGCACTGAGT</td>
<td>GGTTCGCTGTGAGAACC</td>
</tr>
<tr>
<td>Sulfuryl-isomaltase</td>
<td>SI</td>
<td>X36597</td>
<td>CAGGGCTTCTGTTGGTATGC</td>
<td>GGAAGCGAATTTTTTCTTC</td>
</tr>
<tr>
<td>INFRC</td>
<td>-</td>
<td>[ref 18]</td>
<td>GTGGAATAGTTGATGAC</td>
<td>CACCTGTGACACTTCA</td>
</tr>
</tbody>
</table>

Reproducibility and specificity of 3′ amplification of cDNA from single cells

The reproducibility of this technique was assessed when the expression of 20 genes was examined in duplicate on single lymphoblastoma cells. Duplicate PCR assays revealed almost identical expression profiles between assays on a single cell and between different cells (Fig. 3). The expression of 11 housekeeping genes was reproducible in all of these single cell expression profiles, demonstrating consistent amplification of cDNA from the transcripts of these genes. However four genes (JUND, EF1, CDC25B and CD19) were not found to be expressed in every cell tested. Two gene-specific assays, sucrase-isomaltase and CD2, were included as negative controls as they were not expected to be expressed in these cells since their mRNAs have only previously been observed in the gastrointestinal tract (12) and in populations of T-cells (13), respectively. Another set of primers, designed to amplify an intronic sequence from a gene found in the Xq 2.5 region (R.Brooksbank and A.Coffey, personal communication), were used to detect genomic contamination. All three sets of primers have been demonstrated to provide a sensitive assay for their cDNA and genomic DNA targets, respectively (data not shown). Since these assays were negative in all the cells examined (n = 20), and expression of any gene was not detected without prior reverse transcription, false positives arising from amplification of genes not expressed in these cells (SI and CD2) or genomic contamination, do not appear to complicate interpretation of results.

Analysis of complex cell systems: expression profiling of neurons

Having shown that TPEA-PCR permits expression profiling of single cells, we wished to apply it to the analysis of a complex cell system in vivo. Striatal cholinergic interneurons are readily identifiable in rat brain slices due to their large size (> 30 µm diameter) when compared to surrounding cell types, which are predominantly medium spiny neurons (<15 µm diameter). After electrophysiological characterisation of the cells (14), cytoplasmic samples were taken using a patch pipette, reverse transcribed and the cDNA amplified. The expression of a variety of genes of interest was then investigated (two representative expression profiles of cholinergic interneurons are shown in Fig. 4). The expression of four known housekeeping genes was demonstrated, confirming the integrity of sample collection and of the RNA. In addition, the expression of choline acetyltransferase (the acetylcholine synthesising enzyme) was observed in each, thus unequivocally confirming the cholinergic phenotype of the sampled neurons. In order to control against the possibility of contamination by the surrounding population of medium spiny neurons, which are known to express the adenosine A2a receptor (15), we assayed the samples for the presence of the mRNA of glutamic acid decarboxylase (Gad67), which is highly expressed in medium spiny, but not cholinergic neurons. The Gad67 primers amplify from medium spiny neurons (data not shown), and the absence of Gad67 expression in cholinergic interneurons demonstrates that samples had not been contaminated. All neurons tested for the expression of the neurokinin (NK) receptors...
were negative for NK2 and NK3 receptor mRNA, but positive for that of the NK1 receptor. 27% (7 out of 26) of the cholinergic interneurones tested expressed the adenosine A2a receptor.

DISCUSSION

Current methodologies for analysing gene expression in small samples of RNA, particularly from single cells, are severely limited in terms of the number and diversity of genes that it is possible to analyse, and the difficulty of experimental procedures. The methodology outlined here overcomes many of these limitations.

Sequence diversity between genes is at its greatest in the 3′ untranslated region (16) and this region provides the most unique target for gene-specific assays; this is especially important when wishing to differentiate between closely related members of a gene family. Our aim in designing this procedure has therefore been to amplify preferentially this portion of the mRNA sequences. cDNA synthesis by reverse transcriptase is initiated by an anchored oligo-dT priming so that the 3′ region of all genes is represented in the resulting single-stranded cDNA. A 5′-specific heel is incorporated into this primer for use in the subsequent amplification procedure. For second-strand synthesis, it was seen as desirable that ~1 kb from the 3′ end of each gene could be selected and amplified by PCR. Assuming a completely random length of nucleotide sequence, it would be expected that a given five base sequence would appear every 1024 bases (4^5), even though some nucleotide sequences are more common than others (17). A pentameric sequence was selected that would enable the primer to initiate second-strand synthesis in an arbitrary manner within 1 kb of the 3′ end of the mRNA. A search of 30 gene sequences revealed that at least one copy of this five base sequence was present in this region of each gene. 5′ to this, five bases of random sequence (N5) were incorporated in order to stabilize the interaction of the arbitrary pentameric sequence, which in turn was flanked by a specific heel sequence. A similar primer design strategy has been demonstrated to be useful for DOP-PCR primers (18). After a single round of second-strand synthesis, each cDNA contains a specific priming site 5′ and 3′ to the region of interest, thus allowing amplification of the intervening sequence. At present, the exact representation of each mRNA species in the amplified cDNA relative to the starting material is unclear. As far as we have been able to ascertain, however, the majority of the mRNA species represented in the first-strand cDNA pool before amplification, as detected by conventional RT–PCR, were also detectable after amplification.

Problems that can be inherent in the detection of gene expression at the single cell level include the possibility of contamination by other cell types and the interpretation of negative results, which could be due to cell heterogeneity or an artefact of the detection system. In order to minimize these concerns, we chose to develop the system on sorted lymphoblastoma cells. These cells might be expected to be relatively homogeneous, so allowing an assessment of assay reproducibility by comparison of expression profiles between different cells and would also
avoid contamination by other cell types. The use of sorted cells also provides a ready source of single cell material.

Figure 3 clearly demonstrates the reproducibility of the gene-specific assays following amplification from single cells, with each of 11 housekeeping gene assays being detectable in duplicate reactions on each of four lymphoblastoma cells. There was some apparent variability in the expression of four genes (JUND, EF1, CDC25B and CD19) studied in these cells. It is unclear at this stage, however, whether this represents experimental variability, fluctuations in transcriptional activity within these cells or 'real' consistent differences in the expression of these genes in cells that we can only assume to be homogeneous. Such variability in gene expression has been encountered in other cell groups thought to be homogeneous (19). This analysis also demonstrates that the procedure does not lead to false positives due to either over amplification (as no signal was detected for genes SI and CD2 known not to be expressed in these cells) or genomic contamination (as demonstrated by the lack of signal from intronic primers or from the gene-specific primers when the cell contents were not reverse transcribed prior to amplification).

The power of this technique lies in its potential to facilitate expression profiling of cells derived from complex cell populations, even when they form only a small proportion of the population as a whole, and in its ability to detect low abundance transcripts. We therefore investigated the expression of the neurokinin (NK) receptors and the adenosine A2a receptor in single striatal cholinergic interneurons, which constitute a small fraction of the total cellular mass of the striatum. Of the neurokinin receptors, the NK1 receptor is widely accepted as being expressed in these cells (20), so the expression of this gene was examined as an example of an mRNA species expressed at far lower levels than housekeeping genes. In contrast there is considerable controversy as to whether the A2a receptor is expressed in cholinergic interneurons (15,21–23), suggesting that the corresponding mRNA species may be present at low levels as suggested by one in situ hybridisation study (24) or not present at all, as suggested by other studies (25). Expression of the A2a receptor in these cells has important implications for the mechanism of action of adenosine and because of the potential of the A2a receptor to act as a target for novel drugs for the amelioration of Parkinson’s Disease (22). Using a patch pipette it was possible to harvest upwards of an estimated 40% of the cellular contents of these neurons (Fig. 4a). This was then subjected to the cDNA amplification procedure, followed by gene-specific PCR assays. Testing for the expression of four housekeeping genes and the marker enzymes choline acetyltransferase (found only in cholinergic, medium spiny neurons), three tachykinin (NK) receptors and the adenosine A2a receptor was assessed in 26 striatal cholinergic neurons. Two representative neurons are shown; 1 expresses the A2a receptor; 2 does not, while the expression of the other genes tested are the same in both.

![Figure 4. Demonstration of adenosine 2a receptor expression in striatal cholinergic interneurons.](image)

(a) An infrared video image of a rat striatal cholinergic interneuron during electrophysiological characterization and (b) after aspiration of cytoplasm. The expression of four housekeeping genes, the transmitter synthesising enzymes choline acetyltransferase (found only in cholinergic neurons) and glutamic acid decarboxylase (found in GABAergic, medium spiny neurons), three tachykinin (NK) receptors and the adenosine A2a receptor was assessed in 26 striatal cholinergic neurons. Two representative neurons are shown; 1 expresses the A2a receptor; 2 does not, while the expression of the other genes tested are the same in both.
observed previously (24). It is not clear whether the apparent heterogeneity in expression of this receptor is due to differences in the temporal expression of this gene (i.e. that all of these cells possess A2a receptor protein, but only 27% actually express the gene at any one time), or to an absolute difference in gene expression within the striatal cholinergic interneuron population.

An important advantage of TPEA-PCR is the relative ease of the methodology. Other protocols can be time-consuming and complex, involving DNA purification and precipitation from one step to another. The current cDNA amplification technique, however, can be carried out in a single tube with a need for only limited manual intervention. This therefore makes it possible to amplify large numbers of samples relatively easily. The ability to then analyse the expression of many genes of unrelated sequence, both at high and low abundance, in samples taken from as little as a single cell, will potentially allow it to be used in high throughput screening systems.

We have demonstrated that this approach can be used to analyse gene expression in samples from as little as a single cell (Fig. 4) to that contained in 100 cells (Fig. 2). Amplification from a single cell currently provides enough material for ∼40 gene-specific PCR reactions. Whilst this is already an improvement over existing protocols, it should theoretically be possible to improve the efficiency of the TPEA reaction to provide far higher yields of 3′ cDNA product. This would then not only allow the number of gene-specific PCR reactions performed on each sample to be increased, but more importantly could allow the procedure to be linked to other analysis procedures. With improvements to the amplification regime and addition of fluorescent or radioactive nucleotide label to the reaction, it should be possible in the future to analyse the product using array based hybridisation technologies (4) which currently require relatively large amounts of RNA for a single assay. Such developments would potentially allow the expression profiling of hundreds or thousands of genes in samples derived from biopsies or single cells.

In summary, this report describes a rapid, robust and reproducible procedure, called TPEA-PCR, capable of amplifying 3′ fragments of cDNA prior to analysis by other techniques. We believe that this will greatly facilitate the analysis of gene expression in samples obtained from small samples of tissue or single cells. In so doing, it will allow the utilisation of the wealth of new sequence data now available, to further our understanding of disease processes and the cellular physiology of complex tissues.

ACKNOWLEDGEMENTS

We would like to thank Dr David Bentley for all his support and constructive criticism. This work has been funded by the Wellcome Trust and Parke-Davis Neuroscience Research Centre. This technique is the subject of UK Patent application: 98 17055.8.

REFERENCES