Cloning the shared components of complex DNA resources

Anthony J. Brookes*, Euan M. Slorach, Karen E. Morrison, Safia J. Qureshi, Derek Blake, Kay Davies and David J. Porteous

Molecular Genetics Group, Institute of Molecular Medicine, John Radcliffe Hospital, Oxford OX3 9DU, UK
1Neurosciences Group and 2Molecular Genetics Group, Institute of Molecular Medicine, John Radcliffe Hospital, Oxford OX3 9DU, UK

Received July 18, 1994; Revised and Accepted September 6, 1994

The complex and repetitive nature of mammalian genomes limits the ability of conventional molecular techniques to recover sequences of interest. Here we describe a rapid and simple procedure for the direct cloning of sequences which are coincident between DNA mixtures of whole genome complexity. The system, called end ligation coincident sequence cloning (EL-CSC), can enrich coincident DNA by greater than $10^6$-fold and overcomes problems associated with repetitive elements. Applying EL-CSC to various paired DNA resources enables the facile cloning of both genomic markers and novel genes. To demonstrate the power of the method we have i) selectively purified single copy sequences from a complete genome, and ii) isolated gene fragments from 260 kb of cloned genomic DNA.

INTRODUCTION

Standard molecular genetic procedures, such as recombinant library analysis, somatic cell genetics and gel electrophoresis, are based upon the principle of sample fractionation followed by serial analyses in order to identify sequences of interest. Such methods will often only enrich target sequences. Dependent upon prior sequence knowledge, the polymerase chain reaction (PCR) (1) can directly purify specific rare DNAs. Of potentially greater value are strategies for 'difference cloning' and 'coincidence cloning', designed to purify target sequences in the absence of sequence information (2-4). These latter approaches entail the processing of paired DNA mixtures in a manner which recovers only the non-coincident, or the coincident components respectively.

A notable exploitation of the difference cloning concept involved using the phenol enhanced reassociation technique (PERT) (5) to react together normal and deleted X chromosomes in order to isolate new region specific markers which facilitated the cloning of both the Duchenne muscular dystrophy and chronic granulomatous disease genes (6). The construction of subtractive cDNA libraries, genomic subtraction studies (7) and, most recently, representational difference analysis (RDA) (8), are further examples using the difference cloning principle. Of these, RDA is the first generalised protocol that can effectively handle genome complexities, is not confounded by repetitive elements and directly purifies non-coincident DNA. We describe here a coincidence cloning method (end ligation coincident sequence cloning, or EL-CSC) that is fully complementary to RDA.

EL-CSC can be applied to many different problems in genome analysis (summarised in Figure 1) in order to achieve i) the isolation of markers from defined genomic regions, or ii) the purification of genes from genomic DNAs. Three alternative procedures have been described for coincidence cloning of genomic markers. Alu-PCR products (9) in common between pairs of somatic cell hybrids that contain overlapping human components can be selectively recovered (10). However, as initially reported, this is only 10-20 fold better than random clone selection and it is limited to the recovery of human inter-Alu DNA. The more recent 'coincidence painting' (11) and 'Prep-ISH' (12) methods are based upon the physical purification of coincident DNAs by in situ hybridisation to metaphase chromosome spreads. The first of these procedures is limited by the availability of appropriately overlapping abnormal chromosomes suitable for FACS purification, and the second by the need for access to facilities for chromosome microdissection.

Previous coincidence cloning strategies for gene identification (13-16) (we shall refer to these as 'cDNA fishing' methods) use genomic DNA to physically isolate ('fish out'), by base pairing, the genes of a mixed cDNA resource for which they encode. Whilst effective for gene identification, in our experience, these procedures can also isolate transcribed sequences that are related to, but not identical with, components of the genomic DNA under investigation.

EL-CSC is an improved version of a previously described approach to coincident cloning (17), simplified by removal of laborious M13 cloning steps. The resulting method is straightforward, will typically enrich coincident sequences by $10^6$ fold and is not confounded by the presence of repetitive elements. The method can be applied to all standard DNA resources, from simple cloned fragments up to and including mixtures of whole genome complexity. In this report we demonstrate the potential utility of EL-CSC by two investigations; a) a study that isolates unique marker sequences from a complex mammalian genome, and b) an experiment that identifies gene fragments by reacting together genomic DNA and primary cDNA.

RESULTS

End ligation and coincident DNA selection

The EL-CSC procedure (detailed in Figure 2) is based upon the formation of inter-resource duplex DNAs (IRDs) by solution...
hybridisation of two denatured DNA resources. We shall refer to these input DNAs as source DNAs I and II. To drive the kinetics of IRD formation, the simplest of the two resources is always employed as the source I DNA and it is used at the relatively high molar quantity of 1 ng per kb. Therefore, if the source I resource is highly complex it is best converted to a 'representation', either by pooling several thousand random clones or by producing PCR amplicons as previously described (8). The source II DNA, regardless of complexity, may be used directly or similarly simplified if desired.

The isolation of coincident elements is critically dependant upon a brief high temperature end-ligation reaction (Fig. 2, step 4). This ligation reaction underpins the method's high specificity and great enrichment potential. EL-CSC thus recovers IRDs that are both i) sufficiently matched internally to survive stringent washings, and ii) perfectly base-paired at their ends, so permitting ligation to 'capture oligonucleotides'. High copy repeat elements are prevented from forming duplexes by a C<sub>4</sub>1 DNA blocking step, and most low copy repeats will form very poorly matched IRDs because of low intra-family homologies (18,19). Together, these intrinsic selections enable EL-CSC to recover only genuinely coincident molecules.

**Quantifying enrichments**

The enrichment achieved by EL-CSC is quantified in terms of an 'enrichment factor'. This is a measure of the efficiency of recovery of coincident molecules relative to the average efficiency for non-coincident DNAs. Since investigations are always structured with the source II resource as the most complex, then the enrichment factor is related to source II target sequences. It does not consider representation levels in the source I mixture. Using this definition provides the most informative measure of the method's performance. To calculate an enrichment factor, the mass ratio of coincident to non-coincident products is divided by the initial mass ratio of target to non-target source II DNAs. Mass ratios and not percentages must be used in order to avoid inaccuracies when dealing with high levels of product recovery.

**Study 1: whole genome marker isolation**

Isolating region specific genomic markers usually entails the laborious serial examination of numerous clones, many of which originate outside the area of interest. EL-CSC offers a more directed approach, based upon the reaction together of genomic resources that contain overlaps of interest. One could, for example, process pairs of somatic cell hybrids in different rodent
backgrounds that share only a small region of human DNA. Alternatively, the DNA of one hybrid DNA could be reacted with sets of human genomic clones taken from sub-chromosomal libraries or derived by microdissection. For still greater refinement one could directly isolate sequences from hundreds of kb immediately surrounding a single available marker. This would entail running two separate pulse field gel tracks with different restriction enzyme digests of one genome, and then isolating DNA from preparative regions of each that hybridise to the available marker. In a typical case, each of these preparative regions would contain about 1% of the total genome. EL-CSC could be used to isolate the DNA common to these resources. This would constitute 0.01% of the genome (300 kb for human DNA) spanning the marker employed as a probe.

For EL-CSC to be useful for marker isolation it must have an enrichment capability able to purify single copy sequences from whole genome source II mixtures. The complexity of the source I DNA is not the limiting factor as it is always both relatively simple and employed at a standard molarity. We therefore devised a stringent test for this application, employing as the source II DNA two somatic cell hybrid genomes, WJX7.4 and WJX11.2 (20), believed to contain 2 and 15 Mb overlaps respectively, with the region of origin of a series of defined source I clones. The source I clones were 32 large microdissection clones (0.5—2.0 kb size range, some containing repeats) derived from a genomic region of approximately 30 Mb (21). Thus, the prediction would be that 2/30 and 15/30 of the 32 microdissection clones would be shared respectively with the WJX7.4 and WJX11.2 genomes. The challenge for EL-CSC was to identify and recover only these coincident sequences. The design of this experiment both i) enables the efficiency of recovery of source I DNAs to be thoroughly evaluated, and ii) provides a stringent test for the method due to the high complexity of the source II material.

To determine which of the 32 source I clones this experiment could potentially recover, a random 28 of the clones were used as probes upon Southern blots of the hybrid DNAs. This produced 22 interpretable results defining 14 positive assignments to one or other of the hybrid genomes. The remaining six clones gave very weak or repetitive type hybridisations.

After performing EL-CSC as described in Materials and Methods, the PCR products obtained were distinct bands as judged by gel electrophoresis (Fig. 3a). The WJX11.2 resource yielded a greater number of products than the WJX7.4 study consistent with the respective input DNA overlaps. The two sets of reaction products were analysed by direct hybridisation (after isolation by preparative agarose gel electrophoresis as described in Figure 3) onto the 32 input microdissection clones (Fig. 3b). In total, 10 different microdissection clones gave positive signals with one or more of these probes (one being positive with probes from both WJX7.4 and WJX11.2). Of these 11 signals, eight were represented amongst the 14 predetermined assignments. However, this left six potential assignments that had been missed. To investigate why this was so, we probed these six microdissection clones, along with two others as controls (clones which had been shown not to map to the hybrids), onto the EL-CSC PCR products (data not shown). Very weak hybridisation was seen for four of the missed clones while the other two, and the negative controls, gave no detectable signal. PCR biases might explain the poor recovery of these four missed clones and the absence of the other two.

Finally, amongst the 10 clones we did not map conventionally (six unsuccessful when tried and four not tested), three new assignments were identified by EL-CSC, all of which were subsequently confirmed by Southern blotting and hybridisation (example shown in Figure 3c). One of these newly assigned clones belonged to the group of six whose signals did not allow conclusive mapping to the source II hybrids. Previously it gave a very weak signal under standard hybridisation conditions. Its confirmation required both labelling to a higher specific activity and longer exposure times. This demonstrates that EL-CSC can identify markers that a routine hybridisation screening may miss. Furthermore, since all of the EL-CSC derived assignment could be verified, then this experiment did not detectably recover any non-coincident sequences.

Study 2: gene identification
EL-CSC could potentially be employed to identify transcribed sequences in genomic resources by reacting genomic DNA with cloned or uncloned cDNA mixtures. The methods' great

![Figure 3. Application of EL-CSC to marker isolation. The amplified material from the WJX7.4 and WJX11.2 reactions described in the text are shown in (a) next to a 0X174×HaeII marker track (M). The white bars indicate excised regions used to directly probe the 32 source I input clones. (b) Shows the result of one such probing using the largest molecular weight excised region from the WJX11.2 reaction. From left to right the top row contains input clones 1—16 and the bottom row the clones 17—32. (c) Shows an EcoRI genomic Southern blot confirmation of an EL-CSC marker assignment to WJX11.2. Track 'H' is human DNA, tracks WJX7.4 and WJX11.2 are as marked.](https://academic.oup.com/hmg/article-abstract/3/11/2011/569795/Cloning-the-shared-components-of-complex-DNA)
enrichment ability would be expected to recover gene fragments in virtually pure form, thereby minimising the need for extensive product analysis. Furthermore, EL-CSC should be able to recover genes that are expressed at very low levels and, due to ‘breakthrough transcription’ (22), might even be able to recover any gene from any primary cDNA resource.

To evaluate EL-CSC as a tool for gene identification we conducted a study using 260 kb of genomic DNA (pooled cosmid recombinants from chromosome 5) and primary human foetal brain cDNA. A novel brain transcription product (‘gene A’), previously isolated by direct hybridisation of the genomic DNA onto cDNA libraries, provided a positive control for coincident gene recovery. EL-CSC was performed as described in Materials and Methods using complete Sau3AI digestion. The resultant PCR amplified material showed distinct bands with virtually no background smear (Fig. 4a). The total PCR product was cloned into a plasmid vector and analysed as described below. A summary of the results obtained is given in Table 1.

Initially, 23 colonies were chosen at random for content analysis. A direct assessment of insert sizes by gel electrophoresis suggested that 14 clones were probably unique and single pass sequencing of these revealed nine truly distinct products. This analysis identified products containing a fragment from gene A, a rRNA clone and two with internal Alu repeats. The Alu-containing clones were difficult to use as hybridisation probes and so were not investigated further. Also, during subsequent analyses, one of the remaining products that had a particularly small insert (148 bp) did not produce sufficiently clear hybridisation results for reliable interpretation.

This analysis thus identified four potentially genic products that showed no homologies to sequences in available databases. These were tested for coincidence between the input DNAs (Fig. 4b-e). To test for their presence in the input genomic DNA, we hybridised each to the cosmid recombinants and to a chromosome 5 specific somatic cell hybrid mapping panel. All four showed cosmid hybridisation and mapped to chromosome 5. To test for
Table 1. Product summary for gene identification study

<table>
<thead>
<tr>
<th>Clones</th>
<th>Sequence Analysis</th>
<th>Hybridization Analysis</th>
<th>Source I</th>
<th>Source II cDNA (×10^-6)</th>
<th>CHM5 Blot</th>
<th>% of Product (N = 158)</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>1, 2</td>
<td>gene A</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>N/D</td>
<td>5.1</td>
<td>+ve control</td>
</tr>
<tr>
<td>3</td>
<td>rRNA</td>
<td>N/D</td>
<td>N/D</td>
<td>N/D</td>
<td>N/D</td>
<td>? artefact?</td>
<td></td>
</tr>
<tr>
<td>4, 5</td>
<td>Alu</td>
<td>N/D</td>
<td>N/D</td>
<td>N/D</td>
<td>N/D</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Alu</td>
<td>N/D</td>
<td>N/D</td>
<td>N/D</td>
<td>N/D</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>novel</td>
<td>+</td>
<td>0.5</td>
<td>+</td>
<td>N/D</td>
<td>7.6</td>
<td>novel</td>
</tr>
<tr>
<td>9, 10, 11</td>
<td>novel</td>
<td>+</td>
<td>0.5</td>
<td>+</td>
<td>17.1</td>
<td>-</td>
<td>coincident</td>
</tr>
<tr>
<td>12</td>
<td>novel</td>
<td>+</td>
<td>1</td>
<td>+</td>
<td>1.9</td>
<td>-</td>
<td>gene</td>
</tr>
<tr>
<td>13, 14</td>
<td>novel</td>
<td>+</td>
<td>2</td>
<td>+</td>
<td>7.0</td>
<td>-</td>
<td>fragments</td>
</tr>
</tbody>
</table>

Product clones are numbered 1-14 with duplicate clones shown together.
+ and −, positive and negative hybridization respectively; N/D, not done.

their presence in the cDNA, we hybridised each to a Southern blot of this input material including appropriate quantitative controls. All gave positive signals, indicating their generic nature, and all showed low levels of expression. For two cases in which the observed signals mirrored the general area of ethidium bromide fluorescence we confirmed the specificity of the hybridisation by analytical PCR using primers specific to the novel clones (data not shown). Each clone was then additionally hybridised to genomic DNA from multiple mammalian species (zooblot). One clone thus demonstrated evolutionary conservatism.

In order to more accurately determine the representation of each novel product, a larger set of 158 random clones was produced. Hybridisation with total human DNA suggested that 13.9% of these contained medium to high copy repeat sequences. Independent hybridisations with gene A and the four newly identified gene fragments showed that together these accounted for 38.6% of the total cloned material. Experiments are now underway to determine the nature of the gene(s) represented by these novel sequences and also to evaluate further product clones. As part of this work, longer cDNA clones have now been isolated for two of the four novel products (clones '8' and '12') by using them to screen conventional cDNA libraries (data not shown).

DISCUSSION

The selective cloning of sequences that are coincident between complex DNA mixtures holds great promise for both marker and gene isolation. EL-CSC represents a method able to identify coincident DNAs within highly complex DNA resource pairs. To evaluate the potential utility of EL-CSC we have performed two demanding studies attempting to isolate markers and novel gene fragments.

The first study was designed to assay a series of defined markers for their representation in highly complex genomes of somatic cell hybrids. The derived PCR products were entirely genuine coincident DNAs, with a distinct lack of background smear upon ethidium bromide gel analysis. Therefore, given the complexity of the source II DNA (approximately 10^6 fragments) we may deduce that the achieved enrichment factor must be in excess of 10^6. This demonstrates the utility of EL-CSC with mixtures of genome complexity (typically ~10^6 kb). Furthermore, since both input resources contained repetitive DNAs we conclude that the end-ligation aspect of the procedure is effective in blocking the recovery of such 'coincident by default' sequences. Therefore, this 'proof of principle' study suggests that EL-CSC should have a useful role in enabling the rapid isolation of highly focused sets of novel genomic markers.

The second study assesses the ability of EL-CSC to recover transcribed sequences from cloned genomic DNA following its reaction against a complex primary cDNA resource. The number of genes that might be identified by this approach will depend upon the distribution within the genes, of restriction sites that are compatible with the end-ligation step. We would expect the experiment to recover restriction fragments contained within individual exons (the untranscribed region of the 3' exon is on average 0.6 kb and can be considerably larger (24)) as well as fragments spanning closely adjacent exons (given a short internal intron which does not destabilise the IRD). To obtain a theoretical estimate of the efficiency of this strategy we examined the structures of the first alphabetically listed 22 human genes in the GenEMBL database. Parallel studies employing Sau3AI, NlaIII and HpaII would give an average of 6.9 selectable fragments per gene. This figure falls to 5.8 fragments per gene when products that span adjacent exons are excluded (no conclusive evidence for this class of products yet exists). In only one of the 22 tested genes would it be impossible to recover a fragment by EL-CSC, whilst another would yield 24 appropriate fragments. Thus, to optimise EL-CSC based gene hunts one should employ, in parallel, several restriction enzymes that use 4 bp recognition sequences. In this way, experiments are directly focused upon short (~0.3 kb) restriction fragments that will exist within the coding and/or untranslated regions of the majority of genes.

One of the four novel gene fragments identified by the gene finding investigation demonstrated evolutionary conservation, so confirming its generic nature. The other three did not, possibly because they are contained within untranslated generic sequences. Clones containing repeat sequences, as well as the clones not yet examined may also represent additional real products that may or may not be from 5' or 3' untranslated regions. The ribosomal isolate may represent a non-specifically recovered artefact though this has not been fully investigated.

Novel and potentially interesting sequences were identified in the gene finding study by the simple procedure of random clone picking, sizing and sequencing. The fact that this proved to be effective indicates that a major fraction of the recovered clones...
must be coincident gene fragments. Together, the gene A isolate and the four novel genic clones comprise almost 40% of the resultant amplified material, whereas their starting cDNA representations were only of the order of one part in 104. It is difficult to quantify the enrichment level achieved by this study since the product representations in the input cDNA were low (and therefore difficult to accurately determine), and various other potentially real products have yet to be investigated. Furthermore, as noted above, there is likely to be a considerable PCR based variability in the ability of different sequences to be recovered by EL-CSC. Excluding gene A (whose cDNA representation was not determined), we provisionally estimate an enrichment factor of 0.7 × 106 (40 products/60 ‘other’ clones, divided by 10-6).

In summary, this investigation produced a series of clones of which probably the majority represent truly coincident transcribed sequences.

In conclusion, EL-CSC is a general method that we have found to be both effective and robust. Its inherent simplicity and the virtual absence of false products allows coincident DNAs to be quickly isolated. However, as with any PCR based protocol, we would emphasise the need for great care with sample manipulations. This report has considered some of the more obvious applications for EL-CSC. Numerous other uses can be imagined, employing resources such as CpG rich isochore DNA, banks of chromosomally localised clones and established gridded libraries. Applications in which the input DNAs originate from different species can also be envisaged. With further development, EL-CSC may be an effective tool for the rapid localisation of unmapped disease gene loci. This would entail its application to genomes of related affected individuals with the intention of cloning sequences that are ‘identical by descent’ (25,26) and hence flank the region of shared mutation. The simplicity of EL-CSC combined with its wide spectrum of applicability now opens the way to these and other future exploitations of coincidence cloning in complex genome analysis.

MATERIALS AND METHODS

The EL-CSC method

Typically, 20 ng of source 1 DNA is digested to produce Sau3AI compatible ends and ligated at 16°C overnight in a 10 μl volume to 100 ng duplex catch linker 477/479 (5'-CGGAAATTCTAGACTGCAACGCC-3' annexed to kinased 5'-GATCCTGGCTCCTCTAGAATTC-3'). 1 μl of this reaction is PCR amplified with primer 478 (600 ng 5’-biotin-CGGAAATTCTAGACTGCAACGCC-3’). Products are resolved on a 1.2% TAE low melting temperature agarose gel, excised and recovered by agarose digestion (Boehringer Mannheim) and ethanol precipitation. Ten μg aliquots of C51 genomic DNA (species appropriate to the study) are ethanol precipitated and vacuum dried. Source I and source II DNAs are each made to 8.4 μl in H2O and used to separately resuspend the C51 DNAs. 1.8 μl 1 M NaOH is added to each and the samples are placed at 37°C for 5 min. After heating to 50°C, 9.8 μl of 50°C FNET-HC1 mix (6 μl each primer (600 ng for single primer reactions), MgCl2, 2 μl NET, 1.8 μl 1 M HCl) is added to each. Source I and source II samples are then placed at 45°C for 0.5 and 3 h respectively before mixing and incubating overnight (NET comprises 400 mM Tris – HCl, 2.5 M NaCl, 50 mM EDTA, pH 7.8). M280 beads (25 μl) are prepared as recommended by Dynal and incubated at room temperature in 100 μl TEN-S/P (0.1% SDS and 0.5 mg/ml PVP in TEN buffer) for 15 min. This buffer is then replaced with 100 μl TEN-S/P containing 10 μg sonicated salmon sperm DNA and incubated for a further 15 min. After removing this buffer the beads are resuspended in 25 μl TEN-S/P. The reacted DNA mixture is made to 0.1% SDS, 0.5 mg/ml PVP and used to resuspend the beads having first removed the TEN-S/P buffer. The sample is then left at room temperature for 30 min. Washing is performed with six changes of 150 μl TEN at room temperature and four changes of 150 μl 0.1× SSC at 68°C (TEN comprises 100 mM Tris – HCl, 1 mM EDTA, 100 mM NaCl, pH 7.4). The beads are resuspended in 20 μl TEN and heated to 60°C before adding, in 1 μl, 250 ng capture oligonucleotide 732 (5'-GGACGGGTCT-GACACCGGGAGACCTATTTTCAATGGCAGG-3') and 250 ng kinased capture oligonucleotide 735 (5'-GATCCGCGACTCCTGGAATTCACCCGTGCTACCGGAACG-3’). This is then cooled to room temperature over 1 h and washed with four changes 100 μl TEN. A 5 min ligation is performed at 45°C in 20 μl using 10 units of Taq DNA Ligase (Cambio) in recommended buffer and the reaction stopped by adding 1 μl 500 mM EDTA. The sample is then washed with three changes 100 μl TEN at 45°C and six changes of 100 μl 0.1× SSC at 65°C. Product elution is performed by heating the beads to 90°C for 4 min in 10 μl H2O before immobilisation and buffer/eluted DNA recovery. This eluate is then processed through a second magnetic separation to remove all traces of magnetic beads. Coincident DNA is amplified by PCR upon 1 μl aliquots of the eluted DNA using primers 789 (5’-CGTCCGTTGACCGGC-3’) and 596 (5’-GGACGGGTCTGACACCGGGAG-3’). Critical aspects of the procedure are i) care during the multiple sample manipulations to avoid contamination, ii) neutralisation prior to IRD formation (test sample on a pH dipstick), iii) gentleness when washing the Dynal M280 beads, and iv) the maintenance of a constant high temperature during the end-ligation reaction.

Marker isolation study

Source I DNA comprised 100 ng of a mixture of 32 microdissection clones (EcoRI phage recombinants) previously amplified with primers IL (5’-biotin-AATTCGTCGCAGTGAAGCTG-GCAGTGC-3’) and IR (5’-TCTTCCAGGTTAAAA-GCAAAAAGATT-3’). Source II DNAs included 10 μg of representations of the somatic cell hybrid genomes. These were produced by ligating 0.5 to 2.0 kb EcoRI size fractions of each hybrid DNA (isolated by preparative agarose gel electrophoresis) to duplex catch linker 727/731 (5’-GCGAATTCTAGACTGCAACGCC-3’) and annealed to kinased 5’-AATTCTTCACTGATAATTCGACG-3’.

After PCR amplification with primer 727 this was digested to completion with EcoRI. Capture oligonucleotides employed in this study were 647 (5’-GGACGGGTCTGACACCGGGAG-3’), 646 (5’-AATTCGCTTGGACATAACCCGTGCTACCGGAACG-3’) and 645 (5’-AATTCGCTTGGACATAACCCGTGCTACCGGAACG-3’).

Gene identification study

Source I DNA comprised 300 ng of 10 anonymous chromosome 5 human cosmids (including vector sequences). Source II DNA comprised 10 μg of amplified poly-A selected cDNA from human fetal brain. To produce this, Sau3AI digested primary cDNA was ligated to duplex catch linker 727/731 (5’-GCGAATTCTAGACTGCAACGCC-3’) and annealed to kinased 5’-GATACCTGGCTCCTCTAGAATTC-3’ and PCR amplified with primer 727. This was digested to completion with Sau3AI prior to EL-CSC. For cloning, total PCR products were digested to completion with EcoRI, ligated to EcoRI cut and phosphorylated pBluescript plasmid DNA (Stratagene), electro-transformed into XL1-Blue E.coli and grown under ampicillin selection.

PCR amplification

PCRs were performed on a Hybrid Thermal Reactor employing tube temperature control. Reactions were in 50 μl Promega PCR buffer and included 300 ng of each primer (1000 ng for single primer reactions), MgCl2, at 1.5 mM, nucleotides at 200 μM each and 1 U Promega Taq Polymerase. Denaturation times were 30 s for the first cycle and 10 s for subsequent cycles, at 93°C. Annealing times were 30 s at 54°C (primers 478, 727 and IL/IR) or 55°C (primers 789/596). Extension times were 1.5 min for cycles 1–10, 2.5 min for cycles 11–20 and 4 min for cycles 21–30, at 72°C.

Filter hybridisations

Probes were labelled with β32P-CTP using a random priming kit (Boehringer Mannheim). Hybridisations were performed overnight at 68°C in 5 x SSC, 0.1% SDS, 10% dextran sulphate, 5 × Denhardt’s solution and 0.1% pyrophosphate with sonicated salmon sperm competitor DNA at 0.1 mg/ml. Unless otherwise stated, post-hybridisation washes were at 68°C in 0.1 x SSC and 0.1% SDS. Filters were exposed against Kodak XAR5 film or a Molecular Dynamics phosphor-imaging cassette.

ACKNOWLEDGMENTS

The authors acknowledge and thank Drs W. Muir, C. Gosten and A. Weith for the provision of microdissection clones used in this study. This work was supported by the Medical Research Council. AJB and SJQ are funded by the UK Human Genome Mapping Project and KEM is funded by a Wellcome grant.

REFERENCES