# Use of 3′ untranslated sequences of human cDNAs for rapid chromosome assignment and conversion to STSs: implications for an expression map of the genome

Andrea S.Wilcox[1,2], Akbar S.Khan[1,2], Janet A.Hopkins[2] and James M.Sikela[1,2]*
[1]Department of Pharmacology, University of Colorado Health Sciences Center, Denver, CO 80262 and [2]Molecular Biology Laboratory of the Veterans Administration Schizophrenia Center, Denver, CO 80220, USA

Received December 18, 1990; Revised and Accepted March 1, 1991 EMBL accession no. X56960

## ABSTRACT

**A general mapping strategy is described in which the 3′untranslated regions of human cDNAs are used to design PCR primers which will selectively amplify human genomic sequences in a rodent background. When applied to panels of human x hamster somatic cell hybrid DNAs, this approach provides a PCR-based method for rapidly assigning genes to specific chromosomes and chromosomal regions. In addition, it follows from the virtual absence of introns in the 3′untranslated region of vertebrate genes that within this region the cDNA sequences almost always will be identical to those of the genomic DNA and can therefore be used to automatically generate gene-specific sequence-tagged sites (STSs). We have applied this strategy to six human cDNAs and demonstrate that 1) the primers selectively amplify human genomic DNA and 2) the PCR product is of the size predicted from the cDNA. To test this approach further we have utilized it to confirm the known chromosomal location of the retinoblastoma gene. Lastly, we describe how this strategy can readily be applied to unknown human cDNAs, and thereby be integrated into efforts to generate a human STS expression map of the genome. A strategy for production of such a map, using human brain cDNAs as a model, is described.**

## INTRODUCTION

The polymerase chain reaction (PCR) (1) has been used to assign genes to specific chromosomes by taking advantage of the presence of introns (2) or of exon/intron splice sites (3) to generate human-specific PCR products. Because introns display considerably more variability between species than do protein coding regions, PCR products generated from human DNA often can be distinguished from those produced from rodent DNA by using primers that amplify introns. However, these approaches normally require information about the exon/intron organization of the gene that is mapped before useful primers can be designed.

We have developed a general PCR-based gene mapping strategy that requires knowledge of only a few hundred nucleotides of DNA sequence from the 3′untranslated region (3′UT) of the gene of interest, information which is directly available from the cDNA. This method is based on two important features of 3′UTs: 1) they show significant sequence variation between human and rodent homologs and 2) within vertebrate genes 3′UTs are virtually free of introns (4). Because of these properties, PCR primers derived from the 3′UT of human cDNAs very often will selectively amplify a distinct human product from human genomic DNA even within a rodent DNA background, and that product will be of the size predicted from the cDNA sequence. Consequently, it is possible to determine the chromosome from which the cDNA is derived by using such primers for amplification of a panel of human×rodent somatic cell hybrid DNAs.

In addition to being simple and rapid, this strategy automatically generates gene-specific sequence-tagged sites (STSs)(5). Because introns are virtually absent from 3′UT regions, the sequence obtained from the cDNA almost always will be identical to that found in the genomic DNA. Thus, when 3′UT PCR primers are shown to amplify a single product from genomic DNA, and that product is the same size as that predicted from the cDNA, each primer set then defines an STS. In addition, when compared to STSs obtained from regions of the genome for which no function is known, STSs derived from 3′UT regions would have the advantage of being known to be part of expressed genes.

We demonstrate how this approach can be applied to cDNA clones from a human brain library as a model system for the rapid conversion of unknown cDNAs to STSs. We also show how such information can be applied to development of a human brain expression map, which, by providing knowledge of the chromosomal location of the genes expressed in the brain, would be a valuable resource to the study of human neurogenetic disease. Ideally, such a map would be composed of a collection of human brain cDNAs each unique from one another, each with a known

* To whom correspondence should be addressed

chromosomal location, and each in the form of an STS and therefore easily accessible to interested investigators. In addition, each cDNA would be stored in a well of a microtiter plate and therefore have its own address. As part of the development of such a map, we describe a procedure which minimizes the repeated selection of the cDNAs derived from abundant mRNAs, and thus enriches for unique cDNAs. Lastly, generation of a brain expression map would contribute to a stated goal of the Human Genome Project: namely, to generate an STS map of the genome in which STS markers are spaced at approximately 100,000 bp intervals across the genome.

## MATERIALS AND METHODS

### Human and Rodent cDNA Sequences and Comparisons

The human and rodent cDNA sequences of fifteen genes were obtained from Genbank, and separate computer files were created for the protein-coding and 3'UT regions of each cDNA according to the information supplied to Genbank. Both files contained the stop codon to facilitate alignment. The two regions were compared between human and hamster or human and mouse using the N-Align program in PCGene (Intelligenetics). For all comparisons the open gap cost and unit gap cost were set at 10.

### Selection of PCR Primers

PCR primers from the 3'UT sequences of five arbitrarily chosen human cDNAs, one hamster cDNA, and one unknown human brain cDNA were selected using the primer selection program developed by Lowe et al. (6). Selection of primers from human cDNA sequences was done without prior knowledge of corresponding rodent sequences. Parameters chosen included size of the PCR product (50−350 bp), GC content (45−55%), and melting temperature of the oligonucleotide (72.3−82.8°C). Primer sets generating PCR products ranging from 77 to 332 basepairs were chosen so that small changes in fragment size (e.g. due to the presence of introns) could be detected on either 2% agarose or 2% agarose/2% NuSieve (FMC) agarose gels. The only exception to this was the mouse $Ca^{2++}$/calmodulin-dependent protein kinase IV (CaM kinase IV)(7) gene. For this gene primers that were already available were utilized and these amplified a 2.6 kb product from mouse genomic DNA. Studies of numerous vertebrate genes have shown the smallest known intron to be 31 basepairs (4), a size difference which should be evident under the conditions utilized here. To minimize the possibility of amplifying repetitive elements, the DNA sequences within and between primers were compared to the Alu consensus sequence (8) and no primers amplifying such sequences were used. Of the primer pairs that we have designed, Alu-like sequences have been encountered in less than 10% of the cases.

### Synthesis of PCR Primers

Oligonucleotides were synthesized on an Applied Biosystems DNA synthesizer, Model 391A, using the trityl-on mode. After cleavage and deprotection, oligonucleotides were purified on Oligonucleotide Purification Cartridges as directed by the supplier (Applied Biosystems).

### Genomic DNA Samples

Human DNA was obtained from peripheral blood lymphocytes from normal donors using the method of Bell et al. (9). K1 hamster DNA was a gift from Dr. Jingwei Yu (Eleanor Roosevelt Institute for Cancer Research). Mouse DNA was isolated from Institute of Cancer Research mice using a protocol modified from Maniatis et al. (10). Briefly, organs were chopped on dry ice and ground to a fine powder in liquid nitrogen using a chilled mortar and pestle. The powder was incubated for 10 minutes on a rotating wheel at room temperature in 20 ml of TEN9 (10 mM Tris-HCl pH 9.0, 1 mM EDTA, 0.1 M NaCl) containing RNase (2 mg). One ml of proteinase K solution (10 mg/ml) was added and the tube transferred to a 37°C water bath. After two hours, another 1 ml of proteinase K solution was added and the incubation continued overnight. The sample was extracted with phenol, chloroform, and isoamyl alcohol as described for human DNA (9). To improve PCR amplification, these DNAs were sheared by passing them through a 21 gauge needle 20 times. Human×hamster somatic cell hybrid DNAs (PCRable™ DNAs) were purchased from Bios Corporation.

### PCR Amplification of Human and Rodent Genomic DNAs

For PCR amplification of human and rodent genomic DNAs 100 ng of template DNA was used in a 100 μl reaction containing 1.5 mM $MgCl_2$, 200 μM of each dNTP (dATP, dCTP, dGTP, dTTP), 400 ng of each primer, and 2.5 units of Taq polymerase in the reaction buffer (1× is 50 mM KCl, 10 mM Tris-HCl pH 8.3) supplied with the GeneAmp kit (Perkin-Elmer Cetus). Each sample was covered with 50 μl of mineral oil and then amplified in a Perkin-Elmer Cetus thermal cycler. Cycling conditions were: 20 seconds at 72°C, then 35 cycles consisting of 20 seconds at 94°C, 20 seconds at 55°C, and 30 seconds at 72°C (the 72°C extension segment was lengthened by five seconds each cycle). This was followed by a final extension period of 5 minutes at

**Table 1.** Nucleotide sequence identity in protein-coding and 3'untranslated regions of human and rodent cDNAs.

| Genes | Protein-coding (% identity) | 3'-Untranslated (% identity) |
|---|---|---|
| α-myosin heavy chain | 88.7 | 66.7 |
| α-tubulin | 92.6 | 55.7 |
| APRT* | 82.0 | 53.9 |
| β-amyloid* | 89.6 | 66.3 |
| β-2-adrenergic receptor | 87.5 | 80.9 |
| Desmin | 91.2 | 77.8 |
| Dihydrofolate reductase | 87.4 | 55.6 |
| Elongation factor 2 | 87.4 | 66.0 |
| GABA γ2 subunit* | 90.6 | 78.0 |
| HPRT | 92.5 | 82.2 |
| IMP | 89.8 | 80.9 |
| Metallothionein I | 84.9 | 66.4 |
| Ornithine Decarboxylase* | 86.3 | 85.8 |
| Preproglucagon | 88.6 | 75.6 |
| Retinoblastoma* | 88.6 | 78.9 |
| Average | 88.5±2.8 | 71.4±10.6 |

Abbreviations used are: APRT=adenine phosphoribosyltransferase, HPRT=hypoxanthine phosphoribosyltransferase, and IMP=inosine-5'-monophosphate dehydrogenase. Comparisons were made between human and hamster genes except those marked with an asterisk, which were human-mouse comparisons. Sequences were obtained from Genbank or EMBL. Accession numbers for individual human and rodent genes respectively are as follows: α-myosin heavy chain M21664, M12995; α-tubulin K00557, M12252; APRT M16446, M11310; β-amyloid M18734, M18373; β-2-adrenergic receptor M15169, X03804; desmin M26935, K02407; dihydrofolate reductase J00140, K01164; elongation factor 2 M19997, M13708; HPRT M31642, J00060; IMP J04208, J04209; metallothionein V00594, J00061; ornithine decarboxylase M16650, J03733; preproglucagon J04040, J00059; retinoblastoma M15400, M26391. GABA γ2 sequences were from Pritchett et al (27) and Sikela (unpublished data).

72°C. Analysis of products was done by running a 20 μl aliquot of the reactions on 2% agarose gels stained with ethidium bromide.

Amplification of human×hamster hybrid DNAs was done as above, except that the amount of template DNA was 50 ng. In these experiments, 50–90 μl of each PCR reaction was removed, evaporated to approximately 20 μl using a Savant Speed Vac, and analysed on either 2% agarose or 2% agarose/2% NuSieve agarose gels.

### Plaque Hybridization with Human Brain cDNA and Genomic DNA

A human hippocampal cDNA library derived from a 2 year-old female and a human fetal brain cDNA library, both constructed in λZAPII, were obtained from Stratagene. Poly $A^+$ mRNA was extracted by the method of Badley et al. (11) from normal human brains obtained from the National Institutes of Mental Health. First strand cDNA was synthesized by oligo(dT) priming using the Amersham cDNA Synthesis System Plus as directed by the supplier. The resulting cDNA was then precipitated twice with sodium acetate/ethanol.

Screening of the library was done on 15 cm plates of LB agar with 0.6% top agarose (Seakem) using E. coli XL1 Blue cells. Plaque density was approximately 1500 per plate. After overnight incubation at 37°C, plates were refrigerated for two hours. Nylon membranes (Biotrans, ICN) were placed on the plates for five minutes, orientation marks made with India ink, and then the membranes removed from the plates and dried at room temperature for at least 30 minutes. Membranes were soaked for five minutes in a denaturing solution of 1.5 M NaCl and 0.5 M NaOH, then transferred to a neutralizing solution of 3 M sodium acetate, pH 5.5 for two five-minute washes. After drying for 30 minutes at room temperature, blots were baked at 80°C for one hour and stored at 4°C.

Two probes (see below) were labeled with $\alpha^{32}$P-dCTP using random hexanucleotides and the Klenow fragment of DNA polymerase I (Multiprime DNA labeling system, Amersham International). Unincorporated radioactivity was removed by ammonium acetate/ethanol precipitation. The first probe consisted of 50 ng of frontal cortex cDNA having a specific activity of $6\times10^8$ cpm/ug, while the second probe, 50 ng of sheared, denatured human genomic DNA, had a specific activity of $7.5\times10^8$ cpm/μg.

Before probing, membranes were prehybridized at 68°C for one hour in 4× SET (1× is 0.15 M NaCl, 1 mM EDTA, 20 mM Tris-HCl pH 7.8), 10× Denhardt's (1× is 0.02% Ficoll, 0.02% polyvinyl pyrolidone, 0.02% bovine serum albumin), 0.1% SDS, 1.0 ng/ml polyadenylic acid, and 50 μg/ml denatured salmon sperm DNA. The genomic probe was denatured by heating at 95°C for five minutes, added to the appropriate blots, and incubated overnight at 68°C. Where indicated, 100 μg of sheared human genomic DNA was prehybridized with the cDNA probe in a volume of 600 μl in order to block repetitive sequences. This mixture was heated at 95°C for five minutes, then transferred to 37°C for 20 minutes. The probe was added to the appropriate blots and incubated as above. All membranes then were washed in 4× SET, 10× Denhardt's, and 0.1% SDS at 68°C for 20 minutes. This was followed by two 20-minute washes at 68°C in 0.2× SET with 0.1% SDS, and a final four-minute wash in 4× SET at room temperature. Filters were air-dried and exposed to Kodak X-AR film at −70°C with intensifying screens.

### Selection and Storage of cDNAs

Individual plaques selected for inclusion in the collection of cDNAs were removed from the agar plate by coring the plaque with the narrow end of a pasteur pipet and removing only the top agarose layer containing the plaque. Each core was placed into 200 μl of water in a well of a 96-well microtiter plate. Plates were incubated at 37°C for 30 minutes, then the water supernatant was removed and placed in the corresponding well of another microtiter plate. This supernatant was stored at −20°C and used for PCR reactions. 100 μl of SM (10) buffer was added to each of the agarose cores remaining in the original microtiter plate, and these were stored at 4°C. For long-term storage of cores at −70°C, DMSO was added to the SM buffer to give a final concentration of 7% DMSO.

### DNA Sequencing of 3'UT Regions

For double stranded sequencing of the cDNAs, the Bluescript plasmid (Stratagene) containing the cDNA of interest was autoexcised from selected λZAPII cDNA clones according to the protocol recommended by the supplier. Plasmid DNA was

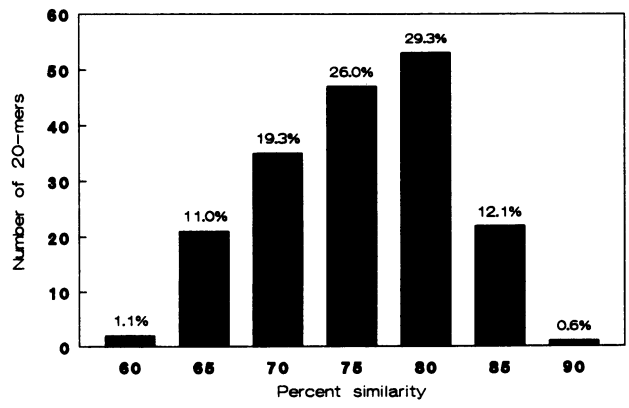


**Figure 1.** Sequence identity of 20-mers from the 3'UT of the human and hamster preproglucagon genes. A total of 181 20-mers from a 200 basepair region of the human 3'UT were compared to the corresponding hamster sequences. Numbers above the bars refer to the percentage of all 20-mers having the indicated similarity.

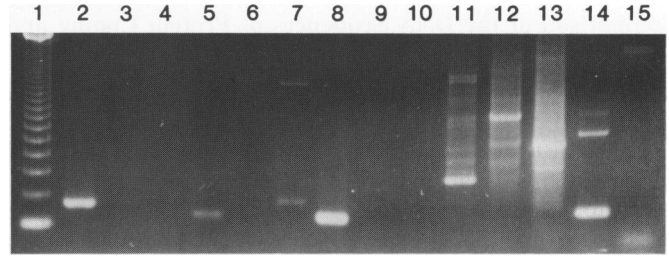


**Figure 2.** Selective PCR Amplification of Human DNAs. Human (lanes 2,5,8 and 11), hamster (lanes 3,6,9 and 12) and mouse (lanes 4,7, 10 and 13) DNAs were PCR amplified with primers derived from several human cDNAs. All human-derived primers were obtained from 3'UT sequences. PCR products were visualized by gel electrophoresis. Lane 1, 123 bp DNA size marker; lanes 2–4, primers were from the HPRT gene; lanes 5–7, primers were from the APRT gene; lanes 8–10, primers were from the dystrophin gene; lanes 11–13, primers were from the β-amyloid gene; lane 14, hamster DNA amplified with primers derived from the hamster HPRT gene; lane 15, mouse DNA amplified with primers from the mouse CaM kinase IV gene.

**Table 2.** PCR Primer Sequences. Primers are shown in the 5′ to 3′ direction. Abbreviations are as in Table 1.

| Gene | Primer Sequence | Product size (basepairs) |
|---|---|---|
| APRT | 5′-CAGGGAAATATCAGCCTTGGG-3′<br>5′-GTGTAATTGGGTTCAGTGTGGC-3′ | 155 |
| β-Amyloid | GCTTCTCTTGCCTAAGTATTCC<br>ATTGAAGACCAGCAGAGCACCC | 332 |
| GABA γ2 subunit | ATGGGTTTTACTGATATGGTTC<br>GATTCAGATACTTATCAACCAC | 104 |
| Human HPRT | ATCTGGAGTCCTATTGACATCG<br>AAACAACAATCCGCCCAAAGGG | 201 |
| Muscular Dystrophy | CAGTAGCAGGACGATGATAGG<br>CTTTTGACTGTGAGAAGAGGGC | 162 |
| Retinoblastoma | ACACAGTATATCCCAAGTGC<br>ACCTTCCCAAATTAAGGAGG | 173 |
| Hamster HPRT | ACCAAGCATTCTAGTTCTGC<br>CAAGGTAAGCGACAATCTATCG | 156 |
| Mouse Cam Kinase IV | ATCTCCCCCTGGTGGGATGAAGTG<br>GGGCGCACAGCACCCTGGCCTCAATCTCAC | ~2,600 |
| cDNA Human Brain 1 | GAAAGTTGGCAAAACTACTGGGC<br>ACCATGTAGCAACAATCTCC | 77 |
| M13 Long Universal | GTAAAACGACGGCCAGTGAATTGT | |
| M13 Long Reverse | GGAAACAGCTATGACCATGATTACG | |

purified by the alkaline lysis method (10). Sequencing of 3′UT regions was carried out according to the protocol described by Khan et al. (12) using an equimolar mixture of three specially designed sequencing primers containing poly(dT)$_{17}$ with either (dA), (dC) or (dG) at the 3′ end. The presence of these three bases at the 3′ end of the primers permits anchoring of the appropriate primer at the upstream end of the poly(A) tail, allowing sequences which lie immediately upstream of the poly(A) tail to be obtained.

## RESULTS

### Comparison of the DNA Sequences of Protein Coding and 3′UT Regions Between Human and Rodent

In order to estimate whether the 3′UT regions of human and rodent differ enough to allow primers to be designed which would amplify a distinct human product, the protein-coding and 3′UT regions of fifteen genes were compared between human and rodent for nucleotide sequence identity (Table 1). Although the similarity of the 3′UT regions varied from 54% to 86% with an average of 71%, they were consistently less alike than the protein-coding regions, which ranged from 85% to 93% with an average value of 89%.

To evaluate further what percentage of randomly-chosen 3′UT human primers might be human-specific (i.e. amplify human but not rodent DNA), we examined a 200 basepair portion of the preproglucagon 3′UT, a region which had an overall sequence identity of 76%. Within a region of this size, it is theoretically possible to generate 181 different oligonucleotides 20 bases long. While most of these 20-mers should have about average similarity when compared to hamster, some will vary more and some less. Those more similar to hamster would be less likely to specifically amplify human DNA. When each 20-mer was compared to the corresponding hamster sequence, only 13% of these theoretical primers showed greater than 80% similarity (Figure 1). The results of the above comparisons suggested that in most cases, primers chosen from 3′UT sequences should amplify a distinct human PCR product even within a rodent DNA background.

### PCR Amplification of Human and Rodent DNAs

PCR primer pairs were selected from the 3′UT sequences of the human HPRT, APRT, dystrophin and β-amyloid cDNAs (Table 2) without prior knowledge of the corresponding rodent sequences. When used for PCR amplification of human genomic DNA, each set of human primers generated a PCR product of the size predicted from the cDNA (Figure 2). When these primers were used with hamster or mouse DNA, either no product was generated, or a band of a different size was seen. However, species-specific primers used with the same hamster or mouse genomic DNAs amplified products of the appropriate size. Amplification of sequences from the dystrophin gene was carried out at an annealing temperature of 60°C because this reduced amplification of nonspecific bands in hamster and mouse DNA.

### Chromosome Assignment by PCR and Automatic Generation of an STS

Using primers to the 3′UT of the human retinoblastoma cDNA, amplification of DNAs from a panel of 25 human×hamster somatic cell hybrids resulted in PCR products from only those hybrids containing human chromosome 13 (Figure 3). This result

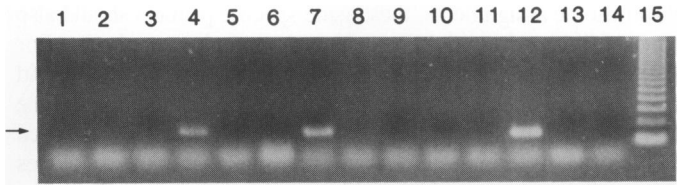


**Figure 3.** Chromosome Assignment of the Human Retinoblastoma Gene. PCR reactions were performed on human×hamster somatic cell hybrid DNAs using primers derived from the 3'UT region of the human retinoblastoma cDNA. Products were electrophoresed as before. Lanes 1–14, somatic cell hybrids 1–14, respectively; lane 15, 123 bp DNA size marker. Human chromosome 13 is only present in hybrids 4, 7 and 12. Arrow denotes PCR product of predicted size.

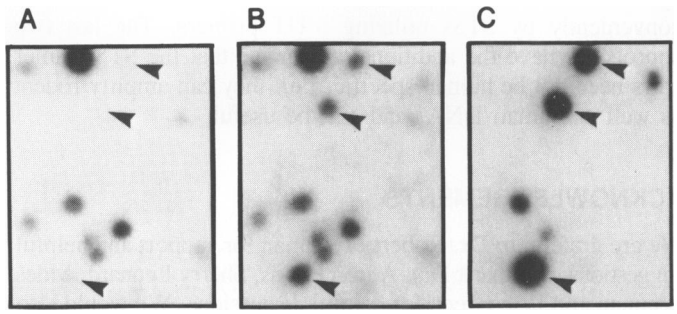


**Figure 4.** Enrichment for Unique cDNAs by Plaque Hybridization. After plating of a human brain cDNA library at low density (~1500 plaques per 15cm plate) three nylon filter lifts (one original and two replica filters) were prepared. The first two filters (A and B) were probed with radioactive human brain cDNA either in the presence (A) or the absence (B) of an excess of cold human genomic DNA. The third filter (C) was probed with radioactive human genomic DNA. Arrows indicate plaques that were positive with the human genomic DNA probe and negative when an excess of cold human genomic DNA was included with the cDNA probe. Such plaques would thus fall into the category of rare mRNAs that contain repetitive sequences and therefore were among those selected.

agrees with previous work which showed by cytogenetic methods that the retinoblastoma gene maps to chromosome 13q14 (13). Because the size of the amplified product was the same as that predicted from the cDNA sequence, we concluded that there were no introns present in the amplified region. Taken together, the chromosomal location, PCR primers, amplification protocol, and sequence information available from the cDNA constitute an STS for the human retinoblastoma gene.

## Application to Generation of a Human STS Expression Map

The generation of a large collection of unique human brain cDNAs is complicated by the fact that some brain messenger RNAs are expressed at much higher levels than others, and thus would be represented in a cDNA library far more frequently. To minimize repeated selection of highly represented cDNAs, we have screened the library with a cDNA probe derived from human brain mRNA so that highly represented and infrequently represented cDNAs could be distinguished. Under these conditions, strongly-hybridizing clones should correspond to abundantly-expressed mRNAs, while weak or non-hybridizing clones should represent rare messages. Selection of clones giving little or no hybridization signal should enrich for unique cDNAs.

Because repetitive sequences present in the cDNA probe could result in strong hybridization signals from clones that actually represent rare mRNAs, we have included an excess of cold human genomic DNA in our hybridizations and demonstrate that such signals can be blocked effectively (Figure 4). Thus this strategy allows us to distinguish those clones corresponding to abundant mRNAs from those representing rare mRNAs which contain repetitive sequences.

```
1  GAAAGTTGGC AAAACTACTG GCTTGGTGGA CGATCAGGAA
41 TAACTTTTCC ACTCTGAGGA GATTGTTGCT ACATGGT 77
```

**Figure 5.** Sequence used to produce an STS for clone HB1. DNA sequence was obtained from the 3'UT region of human brain cDNA HB1 (EMBL accession number X56960). Regions of the sequence used to generate PCR primers are underlined.

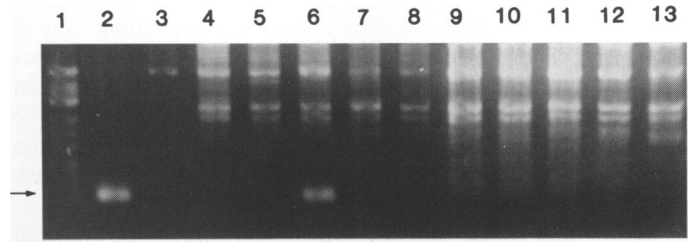


**Figure 6.** Mapping of HB1 to Human Chromosome 15. Primers derived from the 3'UT of human brain cDNA HB1 were used to amplify human, hamster and human×hamster somatic cell hybrid DNAs. PCR products were visualized on a 2% agarose/2% Nusieve gel. Lane 1, φX174/HinfI DNA size marker; lane 2, amplification of human genomic DNA; lane 3, amplification of hamster DNA; lanes 4–13, amplification of somatic cell hybrid DNAs 2–8, 10, 11 and 13. Arrow denotes PCR product of predicted size.

We have used this cDNA selection strategy on two human brain libraries and at present have generated a collection of more than 2,000 clones. Each of these clones is stored in a well of a 96-well microtiter plate and thus has its own address. For each clone two stocks are maintained, one in SM buffer for phage storage and one in water for PCR amplification. Using a metal device having 96 metal prongs, clones from an entire microtiter plate can be plated on an E. coli lawn in one step (14). By slightly offsetting each application, the clones from four microtiter plates can be applied to a single 15 cm plate.

When a commercially available unselected brain cDNA library was plated and screened with total brain cDNA (including cold human genomic DNA) roughly 10–15% of the clones, presumably corresponding to abundant mRNAs, gave hybridization signals above background. In contrast, when the contents of four microtiter plates, thought to represent rare mRNAs, were probed in a similar manner less than 1% of the clones gave signals above background (data not shown), suggesting that the selection strategy had been effective.

In order to demonstrate how the 3'UT/STS mapping strategy and the cDNA selection strategy could be integrated, the mapping approach was applied to an unknown cDNA from the above collection of rare human brain cDNAs. Sequence was obtained from the 3'UT region of the clone and used to design PCR primers (Figure 5). The primers were then used for amplification of human×hamster somatic cell hybrid DNAs (Figure 6). Synteny analysis of the results using 21 human×hamster hybrids indicated the highest concordance frequency (100%) was found with chromosome 15. Assignment of this cDNA to chromosome 15, along with the sequences of the PCR primers and the region they amplify, constitute an STS for this cDNA.

## DISCUSSION

The approach described represents a general strategy by which one can rapidly determine the chromosome location of human genes and at the same time automatically generate gene-specific STSs. Of the seven pairs of 3'UT primers that we selected from human sequences, none of them appear to have produced amplification products from rodent DNA that would have interfered with their use for somatic cell hybrid mapping. Primers which were experimentally shown to amplify only human DNA showed 41% to 80% similarity with corresponding rodent sequences. Those primers which had overall similarity scores near 80%, yet did not amplify rodent DNA, tended to contain mismatches near their 3' ends. Although at present we have applied this approach to a relatively small number of genes, the results of these studies suggest that this strategy should work effectively with a high percentage of human cDNAs.

Along with those already mentioned, there are several additional advantages to using 3'UT sequences as the basis for the identification of large numbers of cDNA clones. First, the use of anchored poly(dT) sequencing primers makes 3'UT sequences immediately upstream of the poly(A) tail directly accessible and particularly adaptable to high throughput automated sequencing strategies (15). Second, obtaining sequence information from the same position of each cDNA, i.e. immediately upstream of the poly(A) region, provides a common reference point for comparison of cDNAs. Generating STSs from cDNAs without regard to such a reference point would likely result in the misidentification of some cDNAs as being unique when in fact the STSs were derived from different parts of the same mRNA. Avoiding such duplication of effort takes on added importance when the objective is to generate as complete a collection of cDNAs as possible in the most efficient manner. While an analysis of known brain mRNAs indicates that alternative polyadenylation sites are utilized in ~20% of brain mRNAs, in roughly half of these (i.e., 10%) the distance between sites is less than 100–200 bp (16). This latter class of mRNAs should be potentially identifiable if 300–400 bp of sequence immediately upstream from the poly(A) tail is read from each cDNA. Thus, possible problems associated with the use of alternative polyadenylation sites may occur with only ~10% of brain mRNAs.

Finally, because 3'UT sequences may show substantial variation between individuals, the gene-specific 3'UT STS amplification products potentially could be useful in the identification of DNA sequence polymorphisms using any of several recently developed PCR-based methods (17–21). Such applications could facilitate integration of a physical STS expression map with the human genetic map.

### Advantages of a Human STS Expression Map

The generation of a human STS expression map, in addition to aiding in the development of an STS map of the genome, would have the potential for significantly improving efforts to identify human genetic disease genes by minimizing the need to employ chromosomal walking methods. For example, if linkage data for a particular disease implicates a specific human chromosomal region, the availability of several cDNAs known to map to that region would provide a number of candidate genes for that disease. These genes then could be studied aggressively.

While we have illustrated how STS primers derived from 3'UT sequences can be used with somatic cell hybrid DNAs for chromosome assignments, these gene-specific primers should also be useful in other higher resolution mapping methods. For example, they can be used to perform PCR on radiation hybrid DNAs to map the cDNA to a particular region of a chromosome (22). In addition, such primers should be applicable to PCR-based methods for assigning cDNAs to particular human YAC clones (23) and thereby provide rapid mapping of these cDNAs to 250–350 kb regions of the genome. Improvements in the sensitivity, speed, and throughput of fluorescence in situ hybridization are allowing large numbers of clones to be mapped with sub-band resolution (24). Examples of mapping 50 clones in one week have been reported (25), and there is evidence that very small probes can be used effectively (26). Thus, it may be possible to use cDNA inserts as probes for fluorescence in situ hybridization. Furthermore, such cDNAs could be identified conveniently by STSs utilizing 3'UT primers. The last two approaches have the additional advantage that the STS primer pairs need not be human-specific, i.e., they can amplify rodent as well as human DNA, and still be useful.

## ACKNOWLEDGEMENTS

We are grateful to Dr. Robert Freedman for support and helpful suggestions. We thank Drs. Adron Harris, Sherry Leonard, Miles Brennan and Eric Green for helpful discussions. We would also like to thank Sherry Leonard for providing human brain mRNA and Cathy Antle for excellent technical assistance. This work is supported by USPHS grants NS27322 and MH44212 and by the Veterans Administration Medical Research Service.

## REFERENCES

1. Saiki,R., Sharf,S., Faloona,F., Mullis,K.B., Horn,G.T., Erlich,H.A. and Arnheim,N. (1985) Science, **230**, 1350–1354.
2. Waseem,A., Gough,A.C., Spurr,N.K. and Lane,E.B. (1990) Genomics, **7**, 188–194.
3. Iggo,R., Gough,A., Xu,W., Lane,D.P. and Spurr,N.K. (1989) Proc. Natl. Acad. Sci. USA, **86**, 6211–6214.
4. Hawkins,J.D. (1988) Nucleic Acids Res., **16**, 9893–9908.
5. Olson,M., Hood,L., Cantor,D. and Botstein,D. (1989) Science **245**, 1434–1435.
6. Lowe,T., Sharefkin,J., Yang,S.Q. and Dieffenbach,C.W. (1990) Nucleic Acids Res., **18**, 1757–1761.
7. Sikela,J.M., Law,M.L., Kao,F.T., Hartz,J.A., Wei,Q. and Hahn, W.E. (1989) Genomics **4**, 21–27.
8. Deininger,P.L., Jolly,D.J., Rubin,D.M., Friedmann,T. and Schmid,C.W. (1981) J. Mol. Biol., **151**, 17–33.
9. Bell,G.I., Karam,J.H. and Rutter,W.J. (1981) Proc. Natl. Acad. Sci. USA, **78**, 5759–5763.
10. Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) Molecular Cloning: A Lab Manual. Cold Spring Harbor University Press, Cold Spring Harbor.
11. Badley,J.E., Bishop,G.A., St.John,T. and Frelinger,J.A. (1988) BioTechniques, **6**, 114–116.
12. Khan,A.S., Wilcox,A.S., Hopkins,J.A. and Sikela,J.M. (1991) Nucleic Acids Research, **19**, 1715.
13. Strong,L.C., Riccardi,V.M., Ferrell,R.E. and Sparkes,R.S. (1981) Science, **213**, 1501–1503.
14. Hochgeschwender,U., Sutcliffe,J.G. and Brennan,M.B. (1989) Proc. Natl. Acad. Sci. USA, **86**, 8482–8486.
15. Wilson,R.K., Yuen,A.S., Clark,S.M., Spence,C., Arakelian,P. and Hood,L. (1988) BioTechniques **6**, 776–787.
16. Sutcliffe,J.G. (1988) Ann. Rev. Neurosci., **11**, 157–198.
17. Orita,M., Iwahana,H., Kanazawa,J., Hayashi,K. and Sekiya,T. (1989) Proc. Natl. Acad. Sci. USA, **86**, 2766–2770.
18. Riesner,D., Steger,G., Zimmat,R., Owens,R.A., Wagenhofer,M., Hillen,W., Vollbach,S. and Henco,K. (1989) Electrophoresis **10**, 377–389.
19. Gibbs,R.A., Nguyen,P., Edwards,A., Civitello,A.B. and Caskey,C.T. (1990) Genomics **7**, 235–244.

20. Hughes,M.R., Malloy,P.J., Kieback,D.G., Kesterson,R.A., Pike,J.W., Feldman,D. and O'Malley,B.W. (1988) Science **242**, 1702–1705.
21. Yandell,D.W. and Dryja,T.P. (1989) Am. J. Hum. Genet. **45**, 547–555.
22. Cox,D.R., Burmeister,M., Price,E.R., Kim,S. and Myers,R.M. (1990) Science, **250**, 245–250.
23. Green,E.D. and Olson,M.V. (1990) Proc. Natl. Acad. Sci. USA, **87**, 1213–1217.
24. Lichter,P., Tang,C.C., Call,K., Hermanson,G., Evans,G.A., Housman,D. and Ward,D.C. (1990) Science, **247**, 64–69.
25. Kievits,T., Dauwerse,J.G., Wiegant,J., Devilee,P., Breuning,M.H., Cornelisse,C.J., van Ommen,G.-J.B. and Pearson,P.L. (1990) Cytogenet Cell Genet **53**, 134–136.
26. Fan,Y.S., Davis,L.M. and Shows,T.B. (1990) Proc. Natl. Acad. Sci. USA **87**, 6223–6227.
27. Pritchett,D.B., Sontheimer,H., Shivers,B.D., Ymer,S., Kettenmann,H., Schofield,P.R. and Seeberg,P.H. (1989) Nature **338**, 582–585.