Nucleic Acids Research, 1994, Vol. 22, No. 15 **3045–3052**

The human Y4 small cytoplasmic RNA gene is controlled by upstream elements and resides on chromosome 7 with all other hY scRNA genes

Richard J.Maraia*, Noriko Sasaki-Tozawa, Claire T.Driscoll, Eric D.Green[1] and Gretchen J.Darlington[2]

Laboratory of Molecular Growth Regulation, National Institute of Child Health and Human Development, National Institutes of Health, 9000 Rockville Pike, Bethesda, MD 20892, [1]Department of Genetics, Washington University School of Medicine, St Louis, MO 63110 and [2]Department of Pathology, Texas Children's Hospital, Houston, TX 77030, USA

Received May 6, 1994; Revised and Accepted June 22, 1994 GenBank accession no. L32608

ABSTRACT

Ro ribonucleoproteins (RNP) constitute a class of evolutionarily conserved small cytoplasmic (sc) RNPs whose functions are unknown. In human cells four distinctive scRNAs designated hY1, hY3, hY4 and hY5 are synthesized by RNA polymerase III (pol III) and accumulate as components of Ro scRNPs. The previously isolated hY1 and hY3 genes contain upstream sequences similar to the class III promoters for U6 and 7SK snRNAs. Additional mammalian Y scRNA genes have been refractory to cloning due to interference from numerous hY-homologous pseudogenes and studies of hY RNA genes have been sparse. Although homologs of hY1 and hY3 RNAs exist in rodent cells, the smaller Y4 and Y5 RNAs do not which has allowed us to localize the hY4 scRNA gene to human chromosome 7 by assaying for its transcript in rodent X human somatic cell hybrids (SCH). A chromosome 7-enriched yeast artificial chromosome (YAC) library was then screened and the authentic hY4 sequence was isolated by strepavidin – biotin-mediated hybrid-selection followed by poly(dA)-tailing and hemispecific PCR. The region upstream of the hY4 sequence contains a TATAAAA motif centered at –26, a candidate proximal sequence element at –63, and three octamer-like sequences located between –260 and –200. hY4 RNA is readily detectable on Northern blots after transient transfection of the hY4 gene into mouse cells but not after transfection of a construct in which the 5′ flanking region was deleted. SCHs and chromosome 7-enriched YACs were used to demonstrate that all four hY RNA genes reside on human chromosome 7.

INTRODUCTION

Four different small cytoplasmic (sc)RNAs designated hY1, hY3, hY4 and hY5 accumulate in human cells as ribonucleoproteins (RNP) associated with the Ro protein (1,2). Although differential expression of hY RNAs has been documented and in red blood cell precursors appears to be determined at the level of transcription the significance and basis of this differential expression is unclear (1–4). *In vivo* Y RNAs are most abundant in heart and brain tissues (see ref. 5), and this pattern is established during fetal development (R.J.M., unpublished). The presence of autoimmune antibodies directed to Ro RNPs is tightly associated with the development of congenital complete heart block in infants born to anti-Ro positive mothers whose autoantibodies cross the placenta (6). Furthermore, it has been reported that these autoantibodies immunolocalize to the affected infants' cardiac conduction system, and in experimental systems inhibit cardiac cell membrane repolarization (refs. 7,8 and refs. therein). Although initially identified as human autoantigens Ro RNPs have since been found in several species including *Xenopus* (3,9–11). The evolutionary conservation of Ro RNPs coupled with their regulated expression and involvement in human pathologic conditions provide ample evidence to indicate an important biological role for Ro RNPs even though no function has yet been ascribed to them (1–4,6,9–11).

The frog and human 60 kDa protein component of Ro RNPs exhibit greater sequence conservation than do the Y RNAs from these species (11). The evolutionary heterogeneity of Y RNAs is further reflected by the fact that different organisms contain different numbers of Y RNAs ranging from two to four (reviewed in ref. 3). Thus, while highly conserved homologs of hY1 and hY3 RNAs exist in several organisms, homologs of the smaller hY4 and hY5 RNAs are less conserved and in fact do not exist in rodents (3,11). Human Y scRNAs are synthesized by RNA

*To whom correspondence should be addressed

polymerase (pol) III (1). While the gene for hY3 scRNA contains consensus sequences corresponding to a TATA box, a proximal element, and an octamer motif similar to the upstream promoters and enhancers which control U6 and 7SK snRNA transcription, the hY3 gene has not been subjected to detailed analyses (2,12–14). Elucidation of the structures of additional hY RNA genes should provide insight into the regulation of hY RNA expression as well as Y RNA gene evolution.

Attempts to clone the human Y4, mouse (m)Y1, and rat (r)Y1 genes yielded sequences which contained multiple mismatches to the corresponding RNA sequence in addition to other characteristics which indicated that the isolated clones represented transposed pseudogenes (3,15–17). Transposed pseudogenes homologous to various small RNAs including Y RNAs are found widely dispersed in eukaryotic genomes. In most cases these defective sequences far outnumber the authentic genes (15,16,18–22; reviewed in ref. 17). Human Y4 scRNA is an homogeneous species as demonstrated by fingerprint and sequence analyses (1,4). This suggests that despite the abundance of Y-homologous sequences in human DNA, hY4 RNA is produced by a single gene (or few genes with identical coding sequences). This fact coupled with the absence of Y4 RNA in rodent cells has allowed us to localize a productive hY4 scRNA gene to human chromosome 7, by assaying for hY4 RNA in a panel of rodent X human somatic cell hybrids. A chromosome 7-enriched yeast artificial chromosome (YAC) library was then screened and yielded several pseudogenes and one authentic hY4 gene which was isolated. Its structure reveals upstream sequence elements similar to those of U6 and 7SK RNA genes as well as the expected pol III termination signal. It is demonstrated that expression of the hY4 RNA gene is dependent on the presence of upstream control elements in transfected tissue culture cells. As such this represents the first demonstration that a Y RNA gene is controlled by upstream elements. It is also demonstrated that the entire family of four hY RNA genes resides on human chromosome 7.

MATERIALS AND METHODS

Cell lines and their human chromosomal content have been described previously (23–25). Maintenance of SCH lines was as recommended by the suppliers. SCH lines A9+2, A9+3, A9+12, A9+15 and GM10481 were kindly provided by O.Pereira-Smith; the A9 derivatives were previously characterized as human monochromosomal hybrids (26). Additional cell lines including GM10791 which contains chromosome 7 as its only human chromosome (27) were obtained from The Coriell Institute for Medical Research (Camden, NJ). Our independent verification of the human chromosomes in hybrid cell lines was done by Giemsa-11 staining which differentially stains rodent and human chromosomes (28), and in some cases also by fluorescent *in situ* hybridization using a human-specific probe. At least fifteen metaphases were analyzed for each hybrid cell line with the exception of GM11010 and GM10478 where twelve and eleven were analyzed, respectively. In all cases the cell lines contained the expected single human chromosome at a frequency within 10% of that published.

Northern blotting was as described previously (25). The oligodeoxynucleotide probes 5′-GCAGTGGGGGGTTGTA-TACCAAC-3′ designated hY4-23, and 5′-AGTGACACTAA-TGTTAATAAGTTC-3′ designated hY4-AS24 are complementary to positions 55–77 and 28–51 of hY4 RNA, respectively (4). The hY5-23 oligo probe 5′-CGGTTGTGGGG-GGAGACAATGTT-3′ is complementary to positions 40–62 of the hY5 RNA sequence (29). Oligo probes were end labelled by polynucleotide kinase and (γ-^{32}P)ATP. Nucleo-cytoplasmic separation was as described (30); for these experiments equal cell equivalents of RNA was analyzed.

Immunoprecipitation was by standard methods (31). Anti-Ro, anti-La and anti-Sm reference sera were obtained from the Centers for Disease Control (Atlanta).

Yeast artificial chromosomes (YACs) containing hY4-homologous sequences were identified by PCR-based screening (32) of a collection of clones higly enriched for human chromosome 7 DNA (33, E.D.G., manuscript in preparation). Specifically, the PCR assay utilized the primers 5′-GGCTGGT-CCGATG-3′ and 5′-AAAAGCCAGTCAAATTTA-3′ (designated hY4-13S and hY4-18AS, respectively) which are directed to the 5′ and 3′ termini of the hY4 RNA sequence (4).

Southern analyses using ^{32}P-labelled oligodeoxynucleotide or plasmid-derived probes were performed according to standard methods as was the preparation of total yeast DNA (33). DNA samples from SCHs as well as from human, mouse and hamster parental cell lines were obtained from the Corriell Institute for Medical Research (Camden) and have been previously characterized with regard to their human chromosomal content (34). Hybridization with oligo probes was done in 6×SSC, 2×Denhardt's, 0.5% SDS, and 100 μg/ml of yeast RNA (30). After oligo probe hybridization washing was done in 2×SSC, 0.1% SDS at moderate stringency (50°C) for 10 minutes or high stringency (60°C) for 20 minutes. The plasmid-derived probe used to localize the hY1/hY3 gene cluster, was generated by random priming in the presence of (α-^{32}P)dCTP of the 0.8 kbp *BamH* I–*Eco* RI fragment isolated from pSW51 (kindly provided by S.Wolin) (2). This was hybridized in 5×SSPE, 0.1% SDS, 5×Denhardt's solution and 100 μg/ml salmon sperm sonicated DNA, at 63°C for 16 hours. Final washing was in 0.2×SSPE, 0.1% SDS for 4 hours at 65°C.

Isolation of the hY4 RNA gene

PCR-mediated amplification of hY4 flanking DNA from YAC yWSS1020 was as follows. First, total genomic DNA from yWSS1020 was digested to completion with restriction endonuclease *Hinf* I (35). This DNA sample was the template for 40 cycles of primer extension-linear amplification using a biotinylated hY4 gene-specific antisense primer: 5′-BBAAAA-GCCAGTCAAATTTAGCAGTGGGGGGT-3′ (B indicates biotin), designated Y4-AS30BI, in a standard single primer thermal cycling reaction. Nonextended primer was removed by gel filtration and the products diluted to 0.5 ml with Tris–EDTA (TE) and incubated with 25 μl of prewashed streptavidin agarose beads (Pierce; ImmunoPure). The beads were sedimented and the supernatant containing unwanted genomic DNA discarded. The beads were washed extensively and drained prior to the addition of 5 μl of 10 mM dATP, 8 μl 5×Terminal deoxynucleotidyl transferase (TdT) buffer (BRL), 24 μl H_2O, and 0.5 μl TdT and incubated at 37°C. After three minutes the reaction was stopped, the 'dT_{17}-adapter primer' and 'adapter primer' described by Frohman (36) were added along with a hY4-specific antisense primer (hY4-23), which was nested with respect to the initial biotinylated primer, and standard PCR was performed. This produced a major product visible by ethidium staining (not shown). To obtain the 3′ flanking region the hY4-13S primer was used with the Alu-specific primer 5′-AAGTCGCGGCCGC-

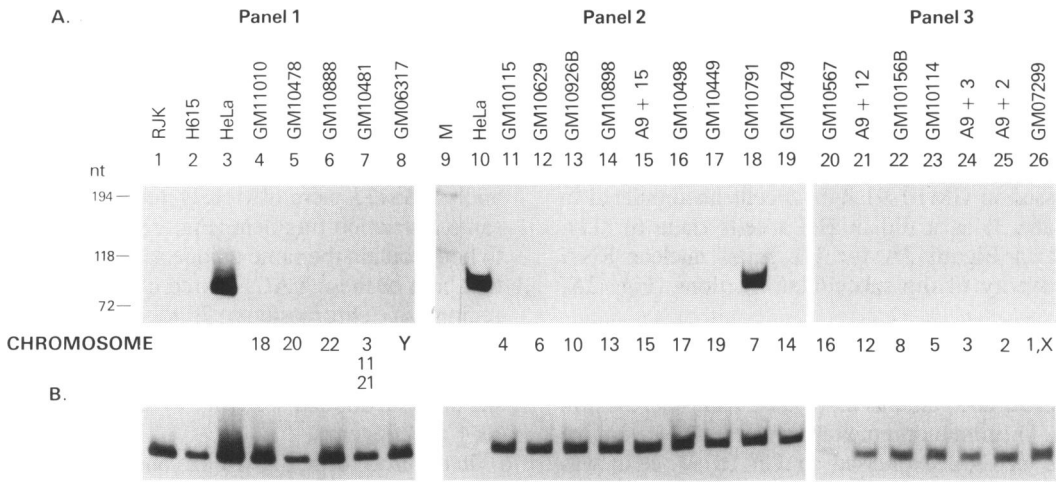

Figure 1. Human chromosome 7 is the only chromosome that expresses hY4 RNA in rodent X human somatic cell hybrids (SCH). (**A**) Total RNA from SCHs were fractionated by denaturing PAGE, transferred to nylon, and probed with ^{32}P-labelled hY4-23 oligodeoxynucleotide. Identities of cell lines are indicated above the lanes and the human chromosome content for each SCH is indicated below. RJK and H615 are Chinese hamster and mouse cell lines, respectively. The human hY4 RNA of 94 nt is detected only in hybrid cell line GM10791, which contains human chromosome 7 as its only human chromosome (25,27,34). Although RNAs from hybrids retaining chromosomes 9 and 16 are not adequately represented on these blots our analyses demonstrated that these chromosomes in SCHs do not produce hY4 RNA. Lane M: co-electrophoresed ^{32}P-labelled denatured HaeIII/ϕx174 size markers are indicated in nucleotides on the left. (**B**) Rehybridization of the blots shown in A with a probe to U1 RNA. The HeLa sample in panel 2 (lane 10) represents cytoplasmic RNA; as expected, it contains little U1 RNA, a nuclear species.

TTGCAGTGAGCCGAGAT-3′ as a hemispecific modification to Alu-PCR (37). After both 5′ and 3′ flanking regions were sequenced by direct PCR sequencing, oligo primers (containing restriction cloning sites) to these distal regions were used to amplify a contiguous fragment from yWSS1020 DNA which contained the internal hY4 gene. This fragment was cloned into the *Bam* HI/*Hind* III sites of pUC 19 and the recombinant designated phY4. The plasmid p5′ΔhY4 was derived from phY4 by attaching *Bam* HI/*Hind* III restriction cloning sites onto the hY4 sequence by PCR, thereby deleting the 5′ flanking sequence up to position minus 6. The integrity of each cloned insert was verified by sequencing recombinant plasmid DNA.

Transient expression of hY4 RNA in NIH 3T3 cells was achieved by plasmid DNA-mediated transfection using Transfectamine (BRL) as described previously (23). Equal amounts of experimental and control plasmids were cotransfected.

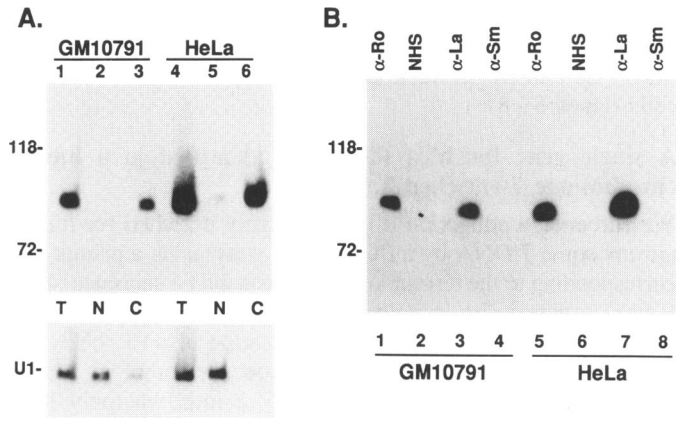

Figure 2. Human chromosome 7 produces hY4 RNA which partitions to the cytoplasm (A) and assembles into Ro scRNPs (B) as examined by Northern blot analysis. (**A**) Subcellular distribution of hY4 RNA in GM10791 cells (lanes 1–3) is comparable to HeLa cells (lanes 4–6). RNA purified from total cell (T, lanes 1 and 4), nuclei (N, lanes 2 and 5), and cytoplasmic (C, lanes 3 and 6) fractions were probed with a hY4-specific 23 nt oligo (upper panel). Size markers are indicated on the left in nucleotides. This blot was then rehybridized with a probe for U1 RNA in order to examine the integrity of subcellular fractionation (lower panel). (**B**) Cytoplasmic extracts of GM10791 (lanes 1–4) and HeLa (lanes 5–8) cells were subjected to immunoprecipitation with the autoimmune anti-sera (anti-Ro, anti-La, anti-Sm) and nonimmune human serum (NHS) as indicated above the lanes. Following RNA purification hY4 RNA was analyzed by Northern blot as above.

RESULTS

An active hY4 scRNA gene resides on human chromosome 7

A well characterized panel of rodent X human somatic cell hybrids comprised of cell lines which contain single human chromosomes was used to identify the chromosome which expresses hY4 scRNA (23,24,25,26,34). Total RNA from these cells was electrophoresed, transferred to a membrane, and hybridized with a hY4-specific 23 nucleotide (nt) probe based on the published hY4 RNA sequence (4). Hybrid cell line GM10791 which contains chromosome 7 as its only human chromosome produced hY4 RNA (Fig. 1A, lane 18). This RNA was not detected in mouse or hamster cells, or in any of the hybrids which contained human chromosomes other than 7. The transcript in GM10791 cells comigrated with HeLa hY4 scRNA (lane 10). To confirm the chromosomal assignment we probed the independent hybrid line GM10482 which contains human chromosome 7 in 70% of the cells (34); as expected, hY4 RNA was readily detectable in these cells (not shown). As a control for RNA integrity and loading the blots shown in Figure 1A were hybridized with an oligodeoxynucleotide probe complementary to U1 small nuclear RNA (Fig. 1B). Thus, chromosome 7 is the only human chromosome that expresses hY4 RNA by this assay.

In order to verify that the transcript produced in GM10791 cells represented authentic hY4 RNA we examined its subcellular localization in Figure 2A and its ability to form a Ro RNP in Figure 2B (1,11). RNA from unfractionated GM10791 cells as well as RNA isolated from nuclear and cytoplasmic fractions were analyzed by Northern blot. Figure 2A shows that as expected hY4 RNA expressed in GM10791 hybrid cells accumulated in the cytoplasm (lane 3) as it did in HeLa cells (lane 6) (11). Probing the blot in Figure 2A for U1 small nuclear RNA confirmed the integrity of the subcellular fractions (Fig. 2A, lower panel).

While the 60 kD Ro protein is specifically associated only with Y RNAs the La protein transiently binds to all nascent transcripts synthesized by pol III and remains associated with Ro RNPs in mammalian cells (1). Immunoprecipitation was used to examine whether the hY4 scRNA expressed in GM10791 cells was assembled into Ro and La RNPs. For this experiment cytoplasmic extracts were incubated with IgG preadsorbed on protein A sepharose; after washing, RNA was purified and examined by Northern blot using the hY4-23 nt probe (Fig. 2B). Anti-Ro (lane 1) and anti-La (lane 3) precipitated hY4 RNA as expected while the control anti-Sm (lane 4) and non-immune (lane 2) sera did not. The fact that hY4 RNA could be incorporated into Ro RNP in rodent X human hybrid cells was not surprising since it was previously shown that human Y3 RNA was assembled into immunoprecipitable Ro RNP using *Xenopus* protein as the source of Ro antigen (11). These experiments demonstrated that hY4 RNA produced in GM10791 cells from human chromosome 7 was assembled into Ro- and La-containing RNPs and was otherwise indistinguishable from hY4 RNA expressed in HeLa cells (lanes 5–8).

A single gene for hY4 RNA was identified in a human chromosome 7-enriched YAC library

We screened a collection of YACs highly enriched for human chromosome 7 DNA by a PCR-based assay using a primer pair corresponding to the termini of the published hY4 sequence (4,33, E.D.G., manuscript in preparation). These PCR primers were designed to specifically avoid amplification of hY1, hY3, and hY5 sequences which share terminal sequence homology and are repetitive in human DNA (15). The 5′ primer was only 13 nt in length and may not detect transposed hY4 sequences which have an incomplete or mutated 5′ terminus (17,21). The nine YAC clones identified by this PCR assay were yWSS132, yWSS133, yWSS182, yWSS1020, yWSS1538, yWSS1694, yWSS2605, yWSS3045, and yWSS3549. The resulting hY4-specific PCR products generated from each YAC were subjected to restriction endonucleases *Acc* I and *Mse* I; the hY4 sequence contains a single site for each of these enzymes (4). Only three of the nine YACs, namely yWSS132, yWSS133 and yWSS1020, were sensitive to both *Acc* I and *Mse* I and therefore represented candidate genes while the others were presumably pseudogenes (data not shown).

Next, Southern blot analyses of the nine YAC-containing clones was performed using two independent non-overlapping oligo probes. The blot was probed first with an oligodeoxynucleotide designated hY4-AS24 which spans the *Mse* I site in hY4 (Fig. 3A). The results indicated that the nine YACs identified by PCR represented multiple different hY4-homologous sequences. Two clones, yWSS3045 and yWSS182 were not reactive with the AS24 probe which spans the *Mse* I recognition site; consistent with this result is the finding that these YACs previously yielded hY4 PCR products which were not recognized by *Mse* I (not shown). YACs yWSS132, yWSS133 and yWSS1020 contained hY4 sequences which were the most reactive with the AS24 probe while the other YACs were less reactive even though all lanes contained comparable amounts of DNA as evidenced by ethidium staining (not shown). The hY4 sequences in YACs yWSS132 and yWSS133 were obviously related since they produced the same restriction fragment (Fig. 3A) and were previously found to both contain the same unique sequence tagged sites indicating that both of these YACs shared overlapping DNA on the same region of chromosome 7 (E.D.G., unpublished). Thus, hybridization and restriction analyses both indicated that YACs 132, 133, and 1020 contained candidate positive hY4 genes while the others exhibited sequence divergence expected of dispersed hY4 pseudogenes.

In order to further scrutinize and discriminate YACs 132 and 133 from YAC 1020 the blot in Fig. 3A was stripped and rehybridized with the hY4-23 oligonucleotide probe used above to assay hY4 RNA on Northern blots. After hybridization this blot was washed at moderate (Fig. 3B, upper) and high stringency (Fig. 3B, lower). Four clones, 2605, 3549, 1694, and 1538 were not detected by this probe which spans the *Acc* I site; consistent with these results is that it was determined that these YACs yielded hY4 PCR products which were not recognized by *Acc* I (not shown). At moderate stringency YACs 132, 133 and 1020 reacted with this probe (Fig. 3B, upper). However, the signal intensity from 132 and 133 decreased to nearly undetectable levels with increased stringency while the signal from YAC 1020 remained (Fig. 3B, lower). Surprisingly, the hY4-23 nt probe detected a band in YAC 182 which corresponded precisely in size to 132 and 133. The hY4 sequence on YAC 182 was previously considered a likely pseudogene based on resistance to *Mse* I (not shown) and nonreactivity with the AS24 oligo probe (Fig. 3A). The 182 hY4-homologous sequence reacts better with the 23 nt probe than with the 24 nt probe, while 132 and 133 exhibit the opposite probe preference. Co-migration of the restriction fragments detected by Southern analysis suggested that the hY4-homologous sequences in YACs 132 and 133 are two representatives of the same locus and represent an allelic polymorphism to the hY4-homologous sequence in YAC 182. This contention is consistent with the fact that diploid human DNA was the original source of YACs 132, 133, and 182. Cumulatively, these results indicate that the hY4 locus occupied in YACs 132, 133, and 182 represents a pseudogene locus which has undergone random sequence drift in two different alleles since neither of these YACs demonstrated as good annealing to the hY4-23 specific probe as did YAC 1020 (Fig. 3B). In any case, the combined results of multiple restriction analyses and filter hybridizations are most consistent with the conclusion that YAC 1020 best represents an authentic hY4 scRNA-encoding sequence by each criteria.

Isolation of the hY4 gene 5′ flanking region by terminal tailing and PCR-amplification of hybrid-selected biotin-captured, endonuclease restricted DNA

After unsuccessful attempts to isolate the region 5′ to the hY4 sequence in YAC yWSS1020 by either ligation-mediated PCR or inverse PCR a novel approach was developed. This procedure is similar to the well-established rapid amplification of cDNA ends (RACE) method of Frohman (36), but confers enhanced specificity to counteract the complexity of genomic DNA as compared to a cDNA. A biotinylated hY4-specific oligonucleotide

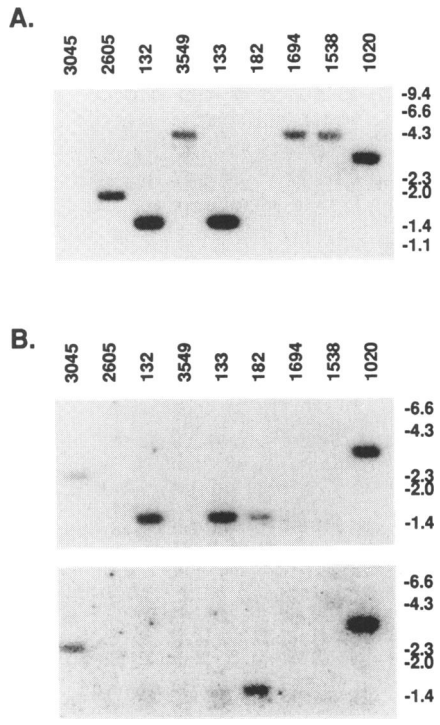

Figure 3. Southern blot analysis of hY4 sequence-containing YAC clones using two independent non-overlapping hY4-specific oligo probes. *Taq* I which does not cut within the hY4 sequence (4) was used to digest total DNA from yeast clones harboring various human chromosome 7-enriched YACs isolated by a PCR assay to detect hY4 sequences (42; see Methods). (**A**) Probe was hY4-AS24 which is complementary to nucleotides 28 through 51 of the hY4 sequence (4) and spans the *Mse* I site at position 35 (see text). Washing was at moderate stringency. Yeast clones are identified above the lanes by their yWSS number. Size markers were coelectrophoresed and are indicated in kilobasepairs on the right. (**B**) Blot in A above was stripped and reprobed with hY4-23 oligo which is complementary to residues 55 through 77 of the hY4 sequence and spans the *Acc* I site at position 60 (see text). Washing was first at moderate stringency (upper panel) then at high stringency (lower panel).

```
-260                -240                -220                -200
CAGGTATTTGCAAAGGACAAGAGGCTTCCATTTGGTAATTTACGTCTATTTCCCATCATGCAACT

                -180                -160                -140
ACATTTATTTTGAATATCAGTTTGCATTAAGCTATTTAAAATTAAAGGAAGATAATCAGATAATG

      -120                -100                -80
TATGGCTTTTGCAGTATTATTTGTAGGGTCATATAATATCAGACTTTTTGGAGAATTCTTAAAAT

      -60                 -40                 -20
AACTCATCCTAACTTATTTAGAGTAGCCACTTCAGAGATTTATAAAAATGAAAGTGAAAGCAGTTT

1                   20                  40                  60
TTCTGGCTGGTCCGATGGTAGTGGGTTATCAGAACTTATTAACATTAGTGTCACTAAAGTTGGTA

                80                  100                 120
TACAACCCCCCACTGCTAAATTTGACTGGCTTTTTACCATTTGATTTTGTATTTCAATCAAACAA

            140                 160                 180
GCTAACTTTACAAAATATTCATGTTGGGGGGAAAGCATTCTTACATAAATGTCCCCCTCCTTTTC

      200                 210                 220
TTTTTAGGGTCTTACTGTGTTGTCCAGGTTGGAGTACAGTAGCACAATCATAGCTCACTGTAGCT
```

Figure 4. Nucleotide sequence of hY4 scRNA gene. The coding region is italicized. The regions of similarity to known promoter elements (12) as discussed in the text are underlined. Note that three different octamer-like sequences exist within the region from −260 to −200 (see Table 2 in ref. 12). The run of 5 (dT) residues corresponding to the pol III terminator is overlined. Only 240 bp of the 3′ flanking region is shown for brevity; the complete sequence up to bp 540 can be obtained through GenBank accession number L32608.

was used for primer extension-linear amplification using endonuclease-restricted yWSS1020 DNA as template. This was followed by immobilization on streptavidin-agarose which removes unwanted genomic DNA template. The retained products were then poly(dA)-tailed and PCR-amplified sequentially on the agarose support. This procedure readily produced a hY4-containing fragment which included 270 bp of 5′ flanking sequence. The hY4 3′ flanking region was isolated by PCR using one hY4-specific primer in conjunction with an Alu interspersed repetitive sequence primer which amplified a ~500 bp fragment from YAC yWSS1020 (37; see Methods). The sequences of the distal termini of both the 5′ and 3′ flanking region products were then determined and primers to these regions were used to PCR-amplify a contiguous fragment of 870 bp containing the hY4 RNA gene from YAC clone yWSS1020 (Fig. 4).

The hY4 scRNA gene contains a TATA box centered at −26, a region between −51 and −66 which is homologous to the proximal sequence elements (PSE) of other small RNA genes, and three short sequences residing between −261 and −200, which are homologous to the octamer-like motifs of certain class II and class III genes (Fig. 4) (see Table 2 in ref. 12). The isolated fragment also contains a run of 5 (dT) residues which corresponds to the termination signal for pol III (overlined in Fig. 4) and the terminal oligo(rU) residues of hY4 scRNA (4). The sequence of the hY4 coding region is in complete agreement with that determined from human reticulocyte hY4 RNA (4) and previous results which demonstrated pppGGC as the 5′ end of hY4 scRNA (5).

Expression of hY4 scRNA by transient transfection requires an intact upstream control region

In order to test whether the cloned sequence represented an authentic and productive hY4 RNA gene the 870 bp fragment containing the hY4 sequence was cloned into pUC19 and introduced into NIH 3T3 cells by cationic liposome-mediated gene transfer. A VAI gene-containing plasmid was cotransfected to control for transfection efficiency and RNA recovery. Total RNA was isolated 42 hours later and examined by Northern blotting. NIH 3T3 cells transfected in duplicate with the hY4 gene efficiently expressed hY4 RNA (Fig. 5, lanes 2 and 3), while cells transfected with vector lacking insert did not (Fig. 5, lane 6). We also tested whether the 5′ flanking region of the hY4 gene was necessary for expression using this assay. The 5′ΔhY4 plasmid was constructed by deleting the 5′ flanking region up to the −6 position. This hY4 construct which contains 6 basepairs upstream of the start site of transcription in addition to the entire hY4 coding sequence and 3′ flanking region, was not expressed in duplicate transfection assays (Fig. 5, lanes 4 and 5), while the control VAI cotransfected plasmid was expressed in the same cells. The results demonstrated that the hY4 RNA gene is dependent on upstream control elements for expression *in vivo*.

All four hY scRNA genes are contained on human chromosome 7

The Northern blots derived from the panel of hybrid cells were used to map the hY5 scRNA gene to chromosome 7 (Fig. 6A). We note that hY5 RNA was expressed at a somewhat lower level in GM10791 cells than in HeLa cells as compared to hY4

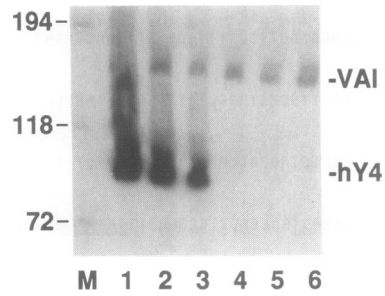

Figure 5. Analysis of hY4 RNA gene expression after transient transfection into NIH 3T3 cells using Lipofectamine (BRL) cationic liposome-mediated DNA transfer as described by the manufacturer. 42 hours after transfection, total nucleic acid was extracted, treated with DNase I, and RNA was isolated and analyzed by Northern blot. All experimental and control plasmids were cotransfected with a plasmid encoding VAI RNA (43). Equal amounts of RNA whose quantity and integrity were determined by polyacrylamide gel electrophoresis and staining were loaded in each lane. Lane 1: non-transfected HeLa cells which serves as a positive control for hY4 RNA; lanes 2 and 3: duplicate transfections with phY4, lanes 4 and 5: duplicate transfections with p5′ΔhY4, lane 6: pUC vector without insert. The blot was first probed and analyzed for hY4 RNA then reprobed for VAI RNA. No VAI-specific RNA was detectable in non-transfected HeLa cells; the smear of signal seen in the HeLa lane is residual from hY4-specific probe.

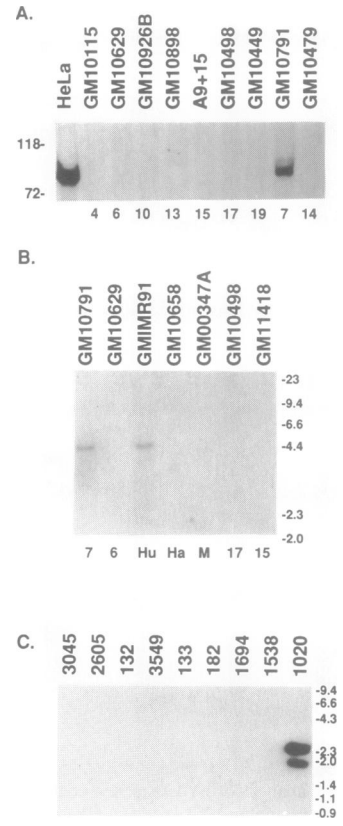

Figure 6. Genes for all four hY RNAs reside on chromosome 7. (**A**) Northern blots shown in Figure 1 were stripped and hybridized with a hY5-specific oligodeoxynucleotide probe. This probe detects hY5 RNA which migrates slightly faster than hY4 RNA and which assembles into Ro and La-containing scRNPs (not shown). GM10791 cells which harbor human chromosome 7 as their only human chromosome express hY5 RNA; since no other cell line in the mapping panel expressed hY5 RNA the full panel of SCHs is not shown here for brevity. Identities of cell lines are indicated above the lanes and the human chromosome content for each SCH is indicated below. (**B**) Southern blot of SCH DNA digested with *Pst* I and hybridized with a 0.8 kbp probe derived from the hY1/hY3 RNA gene locus (2). Identities of cell lines are indicated above the lanes (34). Below the lanes, numbers indicate the human chromosome present in that hybrid cell line, while M, Ha, and Hu indicate mouse, hamster, and human parental DNA respectively. Other hybrid cell lines containing combinations of multiple human chromosomes revealed the ~4.4 kbp band only when chromosome 7 was present, while this band was absent in all hybrid cell lines examined which lacked chromosome 7. (**C**) Southern blot of YAC-derived *Taq* I-digested DNA. The blot shown in Fig. 3 was stripped and rehybridized as in B above. Only YAC 1020 which contains the authentic hY4 RNA gene also contains the hY1/hY3 RNA gene locus represented by the two bands of ~2.4 kb and 1.8 kb (see text).

(compare Fig. 6A with Fig. 1, panel 2). Nonetheless, additional experiments revealed that the hY5 RNA expressed in GM10791 cells was assembled into Ro- and La-containing RNPs and was otherwise indistinguishable from HeLa cell hY5 RNA (not shown). No other human chromosome-containing hybrid in our mapping panel expressed hY5 RNA (not shown). The data demonstrated that an active hY5 RNA gene resides on a single human chromosome which is chromosome 7.

The co-localization of hY4 and hY5 RNA genes to a single chromosome prompted us to examine whether the hY1 and hY3 genes, which were previously found to be adjacent in human DNA, also resided on chromosome 7 (2). A Southern blot of *Pst* I digested DNA from a panel of hybrid cells was hybridized with an 0.8 kbp probe derived from a plasmid containing the hY3 gene (pSW51) (2,34). It was previously shown that a probe from this plasmid detects a fragment of ~4.4 kbp on blots of *Pst* I-digested human DNA (2). Under high stringency conditions, this probe hybridized to a ~4.4 kbp band in somatic cell hybrid GM10791 (Fig. 6B, lane 7) which comigrated with the major band detectable in the total human DNA sample GMIMR91 (lane Hu). Other hybrid cell lines containing combinations of multiple human chromosomes revealed the ~4.4 kbp band only when chromosome 7 was present, while this band was absent in all hybrid cell lines examined which lacked chromosome 7 (not shown). Since GM10791 contains human chromosome 7 as its only human chromosome, and other hybrid cell DNA and rodent parental cell DNA did not contain this fragment (Fig. 6B), these blot results localized the previously described ~4 kb hY1/Y3 gene pair to human chromosome 7 (2,34).

We also probed Southern blots made from yeast clones containing the YACs used to identify the hY4 gene with the hY3 gene-containing plasmid-derived 0.8 kbp probe. This produced a strong signal in YAC yWSS1020 but not in the other YACs tested (Fig. 6C). Two bands of ~2.4 kbp and ~1.8 kbp were detected in *Taq* I-digested yWSS1020 DNA. This pattern is consistent with the *Sal* I–*Taq* I restriction map of the hY3 gene locus which includes a *Taq* I site within the 0.8 kbp probe (*Taq* I cuts within the *Sal* I site; see Fig. 2 of ref. 2). Additional blot hybridizations of this probe to total human DNA and YAC yWSS1020 DNA digested with other restriction enzymes generated the expected pattern of fragments based on the previously reported Southern results and hY1/Y3 restriction map (ref. 2, and data not shown). These data confirm the chromosome 7 localization of the Y1/Y3 gene locus and demonstrate that these genes reside along with hY4 on an interval of DNA of approximately 200 kb, the size of YAC yWSS1020 (E.D.G., unpublished). Furthermore, the lack of hybridzation of this probe to the other YACs present on the blot demonstrated that the

hY4-homologous presumed pseudogenes are not closely linked to the hY1/hY3 gene locus, while the authentic hY4 RNA gene is.

Although attempts to localize a hY5 sequence to YAC yWSS1020 by PCR and Southern analyses have been unsuccessful the combined Northern and Southern blot results demonstrated that the four hY RNA genes are contained on human chromosome 7. Furthermore, hY1, hY3, and hY4 are contained within the 200 kb YAC yWSS1020 while hY5 is not on the region of human chromosome 7 encompassed by this YAC. However, preliminary screening results indicate overlap between a hY4 gene-containing YAC and a YAC which contains a hY5 sequence suggesting that all four hY genes are tightly linked on human chromosome 7.

DISCUSSION

We conclude that a functional hY4 scRNA gene resides on human chromosome 7 and that its expression is controlled by upstream sequences. This work represents the first report that a Y scRNA gene is dependent on 5′ sequences for expression *in vivo*. Transient transfection demonstrated that rodent cells contain the *trans*-acting factors necessary to efficiently express hY4 RNA. Thus, the ability to examine which sequence elements upstream of the hY4 gene control its expression using this system should contribute to our understanding of Y RNA regulation (13,14). The TATA or T/A box, proximal sequence element (PSE), and octamer-like motifs identified in the hY4 gene 5′ flanking region are likely to be involved in pol III transcription of hY4 (12). Although it was noted that hY4 and hY5 RNA levels were found to exceed those of hY1 and hY3 in HeLa and certain other cells, the basis for this differential expression is unknown (1−4). While it is possible that differential stability of the Y RNAs may contribute to the profile of hY RNAs expressed *in vivo*, the study by O'Brien and Harley suggests differential hY transcription in red blood cell precursors (4). It is tempting to speculate that the quality and quantity of the candidate octamer-like motifs within the −260 and −200 region of the hY4 gene might provide this gene greater transcriptional potential than hY1 and hY3 (1−4). The hY3 upstream sequence has been reported to contain a single octamer-like motif ATGCAAAT positioned between −244 and −237 relative to the start site of hY3 transcription (12). Although the presence of a consensus PSE upstream of the hY1 gene has been reported, the presence of an associated octamer-like motif remains to be demonstrated (2,12). The sequences upstream of each of the hY genes will be required to test them for differential transcription, as will a functional analysis of each of the octamer-like sequences of the hY4 gene. TATA elements similar to those 5′ of the hY genes have also been found in each of the four Y RNA genes isolated from *Xenopus* (11). Although this suggests that Y RNA gene structure and presumably regulation, has been conserved through evolution, characterization of more distal upstream sequences of Xenopus Y RNA genes may extend this comparison.

O'Brien and Harley identified hY4 sequences reminiscent of retrotransposed elements derived from RNA polymerase III-transcribed small RNAs (15). Each of the five hY4-homologous presumed pseudogene sequences detected here exhibited less annealing to the oligo probes used as compared to the authentic hY4 RNA gene. Based on the hybridization results with the oligo probes of 23 and 24 nucleotides in length we can estimate a minimal sequence mismatch of one in 24 bases for each of these presumed hY4 pseudogenes, and a presumably higher mismatch for some of them. This approximation is in agreement with results of O'Brien and Harley who identified four and five mismatches (~5%) in each of the three hY4 pseudogenes isolated by stringent colony hybridization (15). Presumably, those represented hY4 pseudogenes most homologous to the authentic hY4 gene and additional more mutated sequences also reside in human DNA. Using the mutation rate of 0.15% per million years dates the insertion time for this subset of hY4 pseudogenes to approximately 30−40 million years ago (38). We detected four distinct loci on chromosome 7 that contain hY4-homologous sequences which appear to be pseudogenes. Assuming random distribution, this predicts about 100 hY4-homologous sequences in the human genome. However, this may not be an accurate estimate because our screening may not have detected hY4-homologous pseudogenes with mutations at their termini (17,21). Regardless of the exact numbers however, our results are consistent with the fact that Y RNA pseudogenes far outnumber the corresponding authentic genes and are abundant in mammalian genomes (3,15,16,39). The dependence of hY4 expression on 5′ flanking control elements explains why the numerous hY4 transposed sequences which are abundant in human DNA are not expressed (15,17,40). It was suggested that hY4 and hY5 RNAs might be products of highly mutated hY sequences derived from the larger hY RNA genes by transposition (15). We note that other than its residence within a generally A+T-rich region, distinctive features of retrotransposed elements such as a discrete 3′ (A)-rich tract and flanking direct repeats were not found in the hY4 scRNA gene locus (17,41). Moreover, the proximity of the hY4 gene to hY1 and hY3 suggests gene duplication as a mechanism of generating this gene family in primates.

Based on the cumulative results presented here it appears that only one authentic hY4 RNA gene exists in human DNA. We can not exclude the formal possibility that another hY4 RNA gene might be expressed from a distinct locus whose expression is not supported in rodent X human hybrid cells. However, the finding that the cloned hY4 RNA gene was expressed with high efficiency in transfected rodent cells argues that if a similar hY4 gene was present on another chromosome it would have been expressed in hybrid cells. Wolin and Steitz reported that hY1 and hY3 genes are single copy in human DNA and are adjacent in the human genome (2). Although the work presented here extends this finding to include the linkage of hY4 and hY5, the exact physical map remains to be determined, as does the sequence and structural relationship of the hY5 gene. Nonetheless, our results are consistent with the possibility that a single gene may encode hY5 RNA as well, since only chromosome 7 expressed this RNA in hybrid cells. At present we do not know where the Y1/Y3 locus is located relative to the Y4 gene within the 200 kb YAC in yWSS1020 or the relationship of either of these loci to the ends of this YAC. However, preliminary data suggests that all four hY genes are tightly linked on human chromosome 7. An important question is whether the functional Y RNA genes in other organisms are as closely linked since this would imply that selective pressure on hY RNA function is dependent on the chromosomal structure of the hY RNA gene family.

ACKNOWLEDGEMENTS

We thank M.Lanigan for secretarial assistance, L.K.Green for contributions to the early phases of hY4 gene cloning, D.Chang and T.Howard for help with transfections, S.Wolin for pSW51

and for sharing information prior to publication, A.Kelekar for useful discussion, T.Kokubo for critical comments on the manuscript and B.Howard and the LMGR/NICHD for support and discussions. E.D.G. is a Lucille P.Markey Scholar and support for this work was also provided in part to E.D.G. by The Lucille P.Markey Charitable Trust and NIH grants P50-HG00201, and to G.J.D. by NIH grants DK45285 and DK44080.

REFERENCES

1. Hendrick, J.P., Wolin, S.L., Rinke, J., Lerner, M.R. and Steitz, J.A. (1981) *Mol. Cell. Biol.*, **1**, 1138–1149.
2. Wolin, S.L. and Steitz, J.A. (1983) *Cell*, **32**, 735–744.
3. Pruijn, G.J., Wingens, P.A., Peters, S.L., Thijssen, J.P. and van Venrooij, W.J. (1993) *Biochim Biophys Acta 1216: 395–401*, **1216**, 395–401.
4. O'Brien, C.A. and Harley, J.B. (1990) *EMBO J.*, **9**, 3683–3689.
5. Wolin, S.L. and Steitz, J.A. (1984) *Proc. Natl. Acad. Sci.*, **81**, 1996–2000.
6. Taylor, P.V., Scott, J.S., Gerlis, L.M., Path, F.R.C., Esscher, E. and Scott, O. (1986) *New England Journal of Medicine*, **315**, 667–672.
7. Deng, J.S., Bair, L.W.J., Shen-Schwarz, S., Ramsey-Goldman, R. and Medsger, T.J. (1987) *Arthritis Rheum*, **30**, 1232–1238.
8. Alexander, E., Buyon, J.P., Provost, T.T. and Guarnieri, T. (1992) *Arthritis & Rheumatism*, **35**, 176–189.
9. Itoh, Y., Kriet, J.D. and Reichlin, M. (1990) *Arthritis & Rheumatism*, **33**, 1815–1821.
10. Lerner, M.R., Boyle, J.A., Hardin, J.A. and Steitz, J.A. (1981) *Science*, **211**, 400–402.
11. O'Brien, C.A., Margelot, K. and Wolin, S.L. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 7250–7254.
12. Hernandez, N. (1992) (eds.), in Transcriptional Regulation. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, pp. 281–313.
13. Meißner, W., Wanandi, I., Carbon, P., Krol, A. and Seifart, K.H. (1994) *Nucl. Acids Res.*, **22**, 553–559.
14. Kelekar, A. and Keene, J.D. (1990) *Mol. Biol. Rep.*, **14**, 173–4.
15. O'Brien, C.A. and Harley, J.B. (1992) *Gene*, **116**, 285–289.
16. Crouch, D. and Liebke, E.H.H. (1989) *Nucl. Acids Res.*, **17**, 4890.
17. Weiner, A.M., Deininger, P.L. and Efstratiadis, A.E. (1986) *Annu. Rev. Biochem.*, **55**, 631–662.
18. Moon, I.S. and Krause, M.O. (1991) *DNA & Cell Biology*, **10**, 23–32.
19. Ullu, E., Esposito, V. and Melli, M. (1982) *J. Mol. Biol.*, **161**, 195–201.
20. Suh, D., Yuan, Y., Henning, D. and Reddy, R. (1989) *Eur. J. Biochem.*, **186**, 221–226.
21. Murphy, S., Altruda, F., Ullu, E., Tripodi, M., Silengo, L. and Melli, M. (1984) *J. Mol. Biol.*, **177**, 575–590.
22. Ullu, E. and Weiner, A.M. (1984) *EMBO J.*, **3**, 3303–3310.
23. Chang, D.Y., Nelson, B., Bilyeu, T., Hsu, K., Darlington, G.J. and Maraia, R.J. (1994) *Mol. Cell Biol.*, **14**, 3949–3959.
24. Driscoll, C.T., Darlington, G.J. and Maraia, R.J. (1994) *Nucl. Acids Res.*, **22**, 722–725.
25. Maraia, R.J., Driscoll, C., Bilyeu, T., Hsu, K. and Darlington, G.J. (1993) *Mol. Cell. Biol.*, **13**, 4233–4241.
26. Ning, Y., Lovell, M., Taylor, L. and Pereira-Smith, O.M. (1992) *Cytogenetics & Cell Genetics*, **60**, 79–80.
27. Jones, N.J., Stewart, S.A. and Thompson, L.H. (1990) *Mutagenesis*, **5**, 15–23.
28. Bobrow, M. and Cross, J. (1974) *Nature*, **251**, 77–79.
29. Kato, N., Hoshino, H. and Harada, F. (1982) *Biochem. Biophys. Res. Commun.*, **108**, 363–370.
30. Maraia, R. (1991) *Nucl. Acids Res.*, **19**, 5695–5702.
31. Steitz, J.A. (1989) In Dahlberg, J.E. and Abelson, J.N. (eds.), Methods in Enzymology **180:**. Academic Press, Inc., San Diego, pp. 468–481.
32. Green, E.D. and Olson, M.V. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 1213–1217.
33. Green, E.D., Idol, J.R., Mohr-Tidwell, R.M., Braden, V.V., Peluso, D.C., Fulton, R.S., Massa, H.F., Magness, C.L., Wilson, A.M., Kimura, J., Weissenbach, J. and Trask, B.J. (1994) *Human Molecular Genetics*, **3**, 489–501.
34. Human Genetic Mutant Cell Repository, N.I.G.M.S. (1992) *National Institute of General Medical Sciences, Bethesda, Md.*, **NIH publication 92-2011**, 579–585.
35. Green, E.D., Mohr, R.M., Idol, J.R., Jones, M., Buckingham, J.M., L., D.L., Moyzis, R.K. and Olson, M.V. (1991) *Genomics*, **11**, 548–564.
36. Frohman, M.A., Dush, M.K. and Martin, G.R. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 8998–9002.
37. Nelson, D.L., Ledbetter, S.A., Corbo, L., Victoria, M.F., Ramirez-Solis, R., Webster, T.B., Ledbetter, D.H. and Caskey, C.T. (1989) *Proc. Natl. Acad. Sci.*, **86**, 6686–6690.
38. Miyamoto, M.M., Slightom, J.L. and Goodman, M. (1987) *Science*, **238**, 369–373.
39. Jurka, J., Smith, T.F. and Labuda, D. (1988) *Nucl. Acids Res.*, **16**, 766.
40. Ullu, E. and Weiner, A.M. (1985) *Nature*, **318**, 371–374.
41. Daniels, G.R. and Deininger, P.L. (1985) *Nucl. Acids Res.*, **13**, 8939–8954.
42. Green, E.D. and Olson, M.V. (1990) *Science*, **250**, 94–98.
43. Maraia, R.J., Kenan, D.J. and Keene, J.D. (1994) *Mol. Cell. Biol.*, **14**, 2147–2158.