

Establishment of a highly sensitive and specific exon-trapping system

(chromosome 11q13/RNA splicing)

MASAAKI HAMAGUCHI*[†], HIROMI SAKAMOTO*, HIROYUKI TSURUTA*, HIROKI SASAKI*, TETSUICHIRO MUTO[†], TAKASHI SUGIMURA*, AND MASAAKI TERADA*

*Genetics Division, National Cancer Center Research Institute, 1-1, Tsukiji 5-chome, Chuo-ku, Tokyo 104, Japan; and [†]Department of Surgery, University of Tokyo, 3-1, Hongo 7-chome, Bunkyo-ku, Tokyo 113, Japan

Contributed by Takashi Sugimura, July 13, 1992

ABSTRACT We have established a highly sensitive and specific exon-trapping system (SETS) with a specific plasmid vector in which an exon in a given DNA segment is identified by its ability to remain as a mature mRNA after splicing. The SETS provides us with the isolation of possible exons rapidly and easily from DNA fragments in chromosomal regions of more than 300 kilobase pairs. Genomic DNA fragments were partially digested and subsequently cloned into plasmid pMHC2, an exon-trapping vector we have constructed. These constructs were transfected into COS-7 cells, and consequent RNA transcripts were spliced in the cells. The resulting mature mRNA was harvested and amplified by using reverse transcription-PCR. Possible exons can be recognized by the sizes of PCR products and cloned into a plasmid vector. The SETS provides a direct means of cloning exons from genomic DNA of more than 300 kilobase pairs within a short period of time. Using this system, we have screened 300-kilobase-pair genomic DNA segments derived from human chromosome 11q13. Human chromosome 11q13 may contain genes responsible for human cancers, because DNA amplification is observed in several malignant tumors. We have successfully identified exon 2 of the *HST1* gene and additional transcribed sequences.

Detailed genetic and physical maps for human chromosomes have been constructed to determine the chromosomal locations of the genes responsible for many human genetic disorders and cancers. The minimal region in which a gene of interest can be found spans several hundred kilobase pairs. Identification and recovery of transcribed sequences from the region of interest are necessary to isolate candidate genes. This strategy has been successful in isolating a number of candidate genes, including Wilms tumor (1–3), neurofibromatosis type 1 (4, 5), and familial adenomatous polyposis (6, 7).

However, available methods for isolation of transcribed sequences from given DNA fragments are inefficient. Those methods most frequently used are interspecies cross-hybridization to search for evolutionarily conserved sequences (3, 8). Other methods include search for open reading frames (9), enhancers (10), promoters (11, 12), or hypomethylated CpG islands (13). But none of these strategies involves direct cloning of transcribed sequences.

Alternative strategies to isolate transcribed sequences involve direct cloning of human transcripts from human-rodent somatic cell hybrids (14, 15). These strategies, however, also have limitations. Possible transcripts derived from human DNA fragments might not be expressed sufficiently enough to be detected in hybrid cells. It is also not easy to establish those hybrid cell lines that contain only the target human DNA fragments.

Two exon-trapping systems based on RNA splicing have been reported and could be important tools for direct cloning of transcribed sequences. However, both of the exon-trapping systems reported previously have disadvantages as described below. A strategy using the retroviral shuttle vector system described by Duyk *et al.* (16) takes much time to complete a round of screening, and the procedure is complicated. The other system, described by Buckler *et al.* (17), worked efficiently; however, its usefulness for identification of exons in a given DNA fragment was still limited because of the following points. First, it was demonstrated that no more than one phage or cosmid can be screened by their system in a single transfection, and no mention was made about the maximum length of the target DNA fragments. The sensitivity of their system would not allow us to apply it to screening for possible exons from a large DNA segment such as is carried in yeast artificial chromosomes (YACs). Second, the combination of vector splice sites and cryptic splice sites of inserts might cause unusual splicing and result in production of false-positive exons. It was reported that three false positives out of four possible exons were generated by Buckler's system. To improve the sensitivity and specificity, we checked genomic DNA sequences of introns and searched for a sequence suitable for a trapping cassette. Check points were the following: first, a pyrimidine tract should be long enough to prevent exon skipping (18). Second, consensus sequences at branch and splice sites should be best fitted to those of small nuclear RNAs involved in splicing. Third, the length of the intron should be longer than 500 base pairs (bp), because a short intron in the trapping cassette makes it difficult to distinguish spliced fragments from unspliced fragments and because most exons are shorter than 500 bp (19). We found suitable sequences for the trapping cassette and have constructed an exon-trapping vector. Here we demonstrate establishment of a sensitive and specific exon-trapping system (SETS) both sensitive and specific enough to carry out large-scale screening of genomic DNA segments for the presence of exons, using this vector. We tested the SETS by isolation of transcribed sequences in overlapping cosmid clones corresponding to human chromosome 11q13 containing *INT2* and *HST1* (20). We have successfully isolated three novel transcribed sequences from the region by this strategy, demonstrating the usefulness of this system, which is a powerful tool for simple screening of megabase genomic DNA segments for the presence of possible exons.

MATERIALS AND METHODS

Construction of the Exon-Trapping Vector. Construction was initiated by PCR amplification of genomic DNA frag-

Abbreviations: SETS, sensitive and specific exon-trapping system; RT, reverse transcription; YAC, yeast artificial chromosome.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

ments for the trapping cassette. The primers (5'-CGT-GAAAGCTTCGAGATGTTCCGA-3' and 5'-CAAGCTTC-TGACGCACATTTATTGCAAGCAA-3') were designed so that PCR products contain part of exon 10, intron 10, and part of exon 11 of the p53 gene, with synthetic *Hind*III recognition sequence indicated by underlining. Genomic DNA of an adult male was amplified by using these primers. PCR products were digested with *Hind*III and partially filled in with dATP and dGTP to produce a fragment of 1 kilobase (kb) with a 5' protruding AG sequence. The parent vector pEUK-C2 (Clontech) was cleaved at a unique *Xba*I restriction site, followed by partial fill-in with dCTP and dTTP to produce a 5' protruding sequence of CT. The PCR products were ligated with the vector to construct the pMHC2 vector (see Fig. 1).

Cell Lines. COS-7 cells (Riken Cell Bank, Tsukuba) were maintained in Dulbecco's modified Eagle's medium containing 10% (vol/vol) heat-inactivated fetal bovine serum. TE-1 cells and TE-10 cells derived from esophageal cancers were grown in RPMI 1640 medium containing 7% heat-inactivated calf serum (21).

Genomic DNA Fragments. Plasmid pLBS6.2 carries 6.2 kilobase pairs (kbp) of genomic sequences of the *HST1* gene (22). Cosmid 35N contains about 40 kbp of uncharacterized genomic DNA segments. Cosmid clones LYI-61, LYH-1, LYH-11, LP-3 [1], LP-9 [2], LP-15 [3], LP-2 [4], and LP-4 [5] contain 30- to 50-kbp genomic DNA segments derived from human chromosome 11q13 (22).

DNA Transfection and Transient Expression. Two micrograms of genomic DNA fragments was partially digested with 0.2 unit of restriction endonuclease *Sau*3AI for 2 hr and fractionated through 1% agarose gels. DNA fragments from 0.6 kbp to 5 kbp were isolated from agarose gels and cloned into a pMHC2 vector that had been cleaved at a unique *Bgl*II recognition site and had been dephosphorylated with calf intestinal alkaline phosphatase (23). Plasmid DNA was prepared by using equilibrium centrifugation in cesium chloride/ethidium bromide gradients. COS-7 cells were transformed with the plasmids by electroporation with a Gene Pulser (Bio-Rad) according to the supplier's protocol. Briefly, 5 million semiconfluent cells were suspended in 0.7 ml of phosphate-buffered saline and placed in a 0.4-cm cuvette, and 10 μ g of a mixture of plasmid DNAs was added. The cells were electroporated at 1.2 kV and 25 μ F. Sixty hours after transformation, poly(A)⁺ RNA was isolated by using FastTrack (Invitrogen, San Diego).

Amplification of Poly(A)⁺ RNA by Reverse Transcription (RT)-PCR. First-strand cDNA was synthesized from poly(A)⁺ RNA by RT from oligo(dT) primers (24) and purified with Bio-Spin 30 columns (Bio-Rad). Primary PCR amplification was performed between the forward primer π A (5'-TGAGGCCTTGGAACTCA-3') and the reverse primer π R (5'-GGAGAATGTCAGTCTGA-3') for 25 cycles. Thirty cycles of secondary PCR amplification were carried out between the forward primer π AB (5'-AGGGATCCCAGGC-TGGGAAGGA-3') and the reverse primer π RB (5'-ACGGATCCTTTTGGACTTCAGGT-3') using 1/100th of the primary PCR products as a template. Both π AB and π RB primers contain a synthetic *Bam*HI recognition sequence indicated by underlining. Positions of the primers are indicated in Fig. 1.

Cloning of Trapped Fragments. The amplified DNA fragments were electrophoresed through polyacrylamide or 4% agarose gels, recovered from the gel (25), digested with *Bam*HI, and cloned into the plasmid vector pBluescript II SK+.

DNA Sequencing. Trapped DNA fragments were sequenced on the Applied Biosystems automated sequencer with *Taq* DNA polymerase in a cycle sequencing protocol (Applied Biosystems) with dideoxynucleotide fluorescent terminators.

Blot Analysis. Poly(A)⁺ RNA was isolated from TE-1 cells or TE-10 cells using FastTrack (Invitrogen), and 2 μ g of poly(A)⁺ RNA was electrophoretically separated in 1% agarose/formaldehyde gels and transferred to Hybond N nylon membranes (Amersham) as recommended by the supplier. Genomic DNA digested with restriction endonucleases was electrophoresed through 1% agarose gels and transferred to Hybond N⁺ membranes (Amersham). To make probes, plasmid DNA with trapped fragments was linearized with *Bss*HIII and transcribed in the presence of ³²P-labeled CTP with phage T3 or T7 RNA polymerases (23). Hybridizations were carried out by standard procedures (23).

Ribonuclease (RNase) Protection Assay. One microgram of pBluescript II SK+ was digested with *Bss*HIII, and labeled RNA probes were synthesized from the linearized plasmids as described above. The RNA probes were hybridized to 1 μ g of poly(A)⁺ RNA from TE-1 cells at 45°C for 16 hr. Unhybridized single-stranded RNA was digested with RNase A and RNase T1 (Ambion, Austin, TX). RNase digestion products were analyzed on denaturing polyacrylamide gels followed by autoradiography.

PCR Amplification of a cDNA Mixture from TE-1 Cells. A cDNA mixture derived from a λ ZAPII library of TE-1 cells containing about 1 million recombinant bacteriophages was amplified by PCR using primers 38A (5'-CATAAGTGC-CGATGGAGTAGGA-3'), 38R (5'-CTCTCCAATTTAGGA-TTCTCA), 40A (5'-TAGCTGACATTCAAAGCCTGGG-3'), and 40R (5'-CTGTGAGGACAGGAGCTGACAGA-3'). These primers were part of trapped fragments and were used for amplification of trapped fragments; 38A and 38R were for MB38 fragments. Primers 40A and 40R were for MB40 fragments (see Table 1).

PCR Amplification. The PCR reaction mixture contained 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM MgCl₂, gelatin at 0.1 mg/ml, dNTPs at 250 nM each, 2 units of *Taq* polymerase (TaKaRa), and 0.5 μ g of primers. PCR amplification of the p53 fragment was carried out in a total volume of 100 μ l for 30 cycles of 94°C for 1 min, 55°C for 30 sec, and 72°C for 2 min. Primary RT-PCR was carried out with 20 cycles followed by 30 cycles of secondary amplification. PCR amplification of the cDNA mixture was performed with 35 cycles of 94°C for 1 min, 60°C for 15 sec, and 72°C for 1 min.

RESULTS

Experimental Strategy. The constructed pMHC2 vector contains the entire intron 10 of the p53 gene, including the long pyrimidine tract and the consensus sequences of the 5' splice site [5'-CAG/GTGAGT-3'], the 3' splice site (5'-AG-3'), and the branch site (5'-TACTCAC-3') (Fig. 1) (26, 34, 35). The pMHC2 vector has a unique *Bgl*II site as a cloning site between the 5' splice site and the branch site (Fig. 1). When an entire exon and the flanking intron derived from the test genomic DNA segments are inserted into the pMHC2 vector in the sense orientation, the transcripts from these plasmids are expected to be processed in the COS-7 cells after transfection, and exons can be identified by RT-PCR. Otherwise, RT-PCR products consist only of the exons of the p53 gene, and a single band of 72 bp is detected by ethidium bromide staining (Fig. 2). Since amplified fragments longer than 72 bp in length contain possible exons and fragments longer than 1 kbp may contain unspliced transcripts, fragments between 90 bp and 900 bp were subcloned into pBluescript II SK+, sequenced, and analyzed for expression.

Trapping of the Known Exon with the SETS. As an initial examination of the system, derivatives of pMHC2 were constructed. pLBS6.2 plasmid, which carries a 6.2-kb DNA fragment including genomic sequences of the *HST1* gene from exon 1 to exon 3, was partially digested with *Sau*3AI, and fragments between 0.6 kbp and 5 kbp in length were

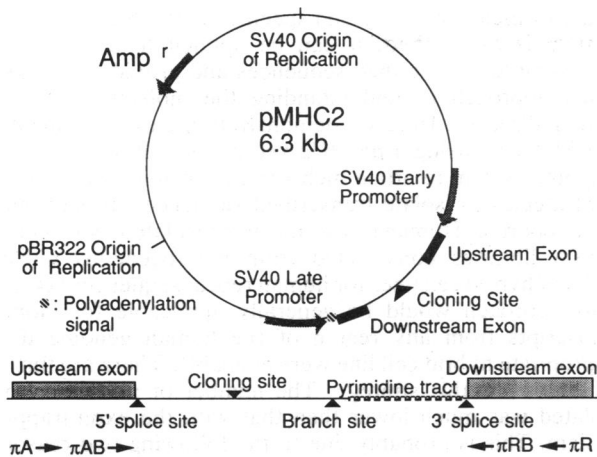


FIG. 1. Structure of the pMHC2 vector, which contains a unique *Bgl* II recognition site as a cloning site. Details of construction of the vector are described in *Materials and Methods*. Amp^r, ampicillin-resistance gene; SV40, simian virus 40.

cloned into the pMHC2 vector (pMHC/HST). The plasmids, pMHC2/HST, were transfected into COS-7 cells. Sixty hours after transfection, poly(A)⁺ RNA was extracted and amplified by RT-PCR as described in *Materials and Methods*. PCR products were separated through polyacrylamide gels, and two bands of 70 bp and 170 bp were visualized by ethidium bromide staining. DNA fragments between 90 bp and 900 bp were recovered from the gel, inserted into pBluescript II SK+, and sequenced. Four out of four plasmids analyzed contained exon 2 of the *HST1* gene, and the exon was trapped precisely between exon 10 and exon 11 of the p53 gene (data not shown).

Estimation of the Sensitivity of the SETS. Cosmid 35N contains genomic DNA segments of 40 kbp and provides no possible exons by the SETS. To estimate the sensitivity of the SETS, 5 μg of 35N and 100 ng of pLBS6.2 were mixed and used for the SETS. pMHC/H35N was constructed by ligation of pMHC2 with a mixture of 35N and pLBS6.2 DNA that had been partially digested with *Sau*3AI. pMHC/H35N was transfected into COS-7 cells and poly(A)⁺ RNA was har-

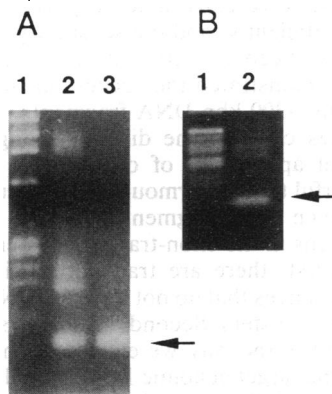


FIG. 2. Electrophoresis of RT-PCR products. The arrows indicate 72-bp bands. (A) Electrophoresis through a 1.5% agarose gel. Lane 1, 200 ng of *Hind*III-digested λ DNA and 200 ng of *Hae* III-digested φX174 DNA as size markers. Lane 2, products of poly(A)⁺ RNA harvested from COS-7 cells 60 hr after transfection with pMHC2/11q13. Lane 3, products from poly(A)⁺ RNA harvested from COS-7 cells 60 hr after transfection with pMHC2. Details of construction of pMHC2/11q13 are described in *Results*. (B) Electrophoresis through a 4.5% agarose gel. Lane 1, 300 ng of *Msp* I-digested pBR322 DNA as size markers. Lane 2, products of poly(A)⁺ RNA harvested from COS-7 cells 60 hr after transfection with pMHC2/11q13.

vested. RT-PCR amplification was carried out, and the PCR products were separated through 4% agarose gels and analyzed by Southern blot analysis using ³²P-labeled cDNA of the *HST1* gene as a probe. As expected, the probe detected a single band of 160 bp (data not shown). The exon of the *HST1* gene was successfully isolated by the SETS. Because the pLBS6.2 clone contains a 6.2-kbp DNA segment and 1/50th of the transfected DNA molecules was derived from the pLBS6.2 clone, the results indicate that by this system possible exons can be identified and recovered from genomic DNA segments with a length of 300 kbp.

Application of the SETS to Human Chromosome 11q13. Eight cosmid clones derived from 11q13, LYI-61, LYH-1, LYH-11, LP-3 [1], LP-9 [2], LP-15 [3], LP-2 [4], and LP-4 [5], contain 30–50 kbp of genomic DNA sequences in each clone. A mixture of these cosmids was used for screening by the SETS. pMHC2/11q13 was constructed as follows: 250 ng each of the eight cosmid clones was partially digested with *Sau*3AI. DNA fragments from 0.6 kbp to 5 kbp were inserted into the pMHC2 vector. Ten micrograms of the plasmids was transfected to COS-7 cells, and all the procedures for exon trapping were carried out as described above. Several fragments >72 bp were detected by ethidium bromide staining (Fig. 2A, lane 2), whereas lane 3 contained only a single band of 72 bp. This demonstrates the successful amplification of possible exons. Electrophoresis through a 4.5% agarose gel demonstrated that only several bands can be visualized, indicating that random amplification did not occur (Fig. 2B). Fragments amplified by RT-PCR were cloned into pBlue-script II SK+, and thousands of transformed plasmid clones were obtained. The inserts of the plasmids are assorted into several groups (Fig. 3), assuring that possible exons should be placed in one of several classifications. Sequence analysis of 35 clones revealed that these clones were assorted into eight independent groups. Seven groups contained exon 10 and exon 11 of the p53 gene with trapped fragments, and the other contained only the p53 gene. Seven trapped fragments were possible exons and examined further. One fragment was exon 2 of the *HST1* gene. To determine the derivation of the other possible exons, each fragment was hybridized with the cosmid panels (data not shown). All fragments originated from the cosmid clones on 11q13 (Table 1).

Characterization of the Six Possible Exons. Interspecies cross-hybridization demonstrated that none of the possible exons except exon 2 of the *HST1* gene carries evolutionarily conserved sequences (data not shown). A screening of the GenBank data base (May 1992) showed no significant similarity among the sequences of these six possible exons and previously reported sequences. All trapped fragments were labeled by *in vitro* transcription and hybridized to poly(A)⁺ RNA of TE-1. The MB38 fragment detects transcripts of 2.7 kb and 1.9 kb, and the MB40 detects transcripts of 2.7 kb and



FIG. 3. Electrophoretical analysis of possible exons. Lane 1, 500 ng of *Msp* I-digested pBR322 DNA. The other lanes contain DNA fragments from plasmids in which trapped possible exons were inserted. Each plasmid was digested with *Bam*HI and electrophoresed through a 3% agarose gel.

Table 1. Characterization of trapped fragments

Clone	Origin	Length, bp	Expression in TE-1 cells
MB14	LYH-1, LYH-11	104*	-
MB15	LYI-61	171	-
MB22	LYH-1	205	-
MB36	LYI-61, LYH-1	112	-
MB38	LYH-11, LP-3 [1]	144	+
MB40	LP-9 [2], LP-15 [3]	173	+
MB42	LP-3 [1]	114	+

*Exon 2 of the *HST1* gene.

1.9 kb. MB42 detects transcripts of 5.2 kb, 3 kb, and 1.1 kb (Fig. 4A). Ribonuclease protection assays demonstrated that the entire sequences of clones MB38, MB40, and MB42 should be exons (Fig. 4B). PCR amplification of the cDNA mixture revealed that the cDNA library contains cDNA clones that are detectable by the MB38 and the MB40 fragments (Fig. 4C). The MB15, MB22, and MB36 fragments did not hybridize to poly(A)⁺ RNA of TE-1.

DISCUSSION

The recent development of reverse genetics provides us with high-resolution maps and YAC libraries of a number of regions in which genes responsible for human diseases can be found. To identify the genes responsible for human diseases, construction of a detailed transcript map is indispensable. A transcript map should be made after subcloning YAC clones into cosmid vectors followed by identification of exons. This step to identify transcripts is essential. In addition, presently available methods for this purpose are inefficient. The exon-trapping system has become an important technique for positional cloning (16, 17). In spite of such successes, the applications of exon-trapping systems were restricted be-

cause of their sensitivity and specificity. If a highly efficient system is established, it can be applied to a large-scale screening for transcribed sequences and provide a sophisticated approach to understanding the molecular basis of human diseases. Here we demonstrate the establishment of the SETS by using a pMHC2 vector that allows the exon-trapping system to attain high sensitivity and strictness.

Strategies to isolate transcribed sequences directly have been reported. Human transcripts were isolated by screening cDNA libraries constructed from heterogeneous nuclear RNA of hybrid cell lines for human repeat sequences (14, 15). This approach would be generally applicable for cloning transcripts from any region of the human genome if an appropriate hybrid cell line were available. There are further limitations to this strategy. The number of possible exons isolated was much lower than that with the exon-trapping system. This is probably due to the following two reasons: first, not all immature transcripts have *Alu* repeats in their introns. Second, all human genes on the chromosomal segments contained in the hybrid cells would not be expressed as much as expected.

Alternative strategies based on RNA splicing have been reported, including a retrovirus system. One strategy using a retroviral shuttle vector system was tedious and inefficient (16). Another exon-trapping system using a pSPL1 plasmid, which contains the human immunodeficiency virus 1 *tat* intron, demonstrated the recovery of an exon of the DNA excision repair gene *ERCC1* from a 20-kbp genomic DNA fragment with an incidence of false positives being 3 times higher than that of true exons (17). Such sensitivity or specificity is not sufficient for application of the exon-trapping system to large-scale screening of uncharacterized DNA fragments.

The step to identify and isolate transcribed sequences still remains a crucial technical problem. As the exon-trapping systems depend only on the structure of the introns, the sequences of the trapping cassette have significant importance. We screened the GenBank data base to find the best sequence for the trapping cassette. As the sequences of the branch sites and the polypyrimidine tract of the inserted DNA have a wide variety, the sequence of the 5' splice site should be best for the recognition of 5' splice site (27). We chose intron 10 of the *p53* gene according to the above requirements and have constructed a pMHC2 vector. This vector worked efficiently, and the sensitivity and specificity of the SETS have become sufficient to screen a large DNA fragment. We demonstrated the successful isolation of possible exons from >300-kbp DNA fragments. This indicates that YAC clones can become direct starting materials of SETS. Potential application of our system to YAC may provide a powerful tool to surmount the difficulty of cloning genes from a given DNA fragment.

Some limitations in the exon-trapping system still need to be improved. First, there are transcripts without introns. Transcribed sequences that do not undergo RNA splicing are not isolated by this system. Second, some types of exons may not be processed in the way we expected. In other words, some exons in the target genomic DNA would be eliminated by splicing. A third limitation is the production of false-positive exons due to activation of cryptic splice sites, although this is not a major problem, because, first, at least four fragments out of seven possible exons identified by SETS were actually present as transcripts by RNA blot analysis or RNase protection assay. Further characterization is required to determine whether the other three fragments are true exons or not. Notwithstanding the above limitations, the SETS is worthy of note.

We applied this exon-trapping system to the 11q13 region. DNA amplification of this region has been reported in a number of human cancers. We previously reported coampli-

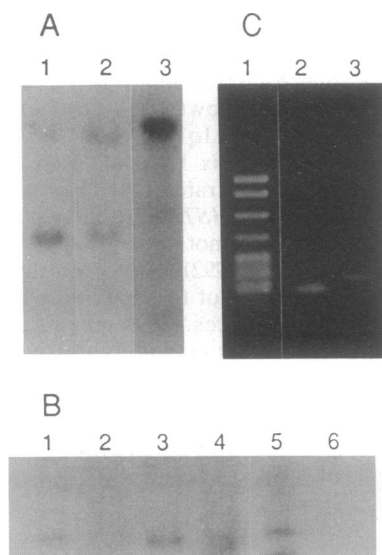


FIG. 4. Characterization of the possible exons. (A) RNA blot analysis of TE-1 using possible exons. Each lane contains 2 μ g of poly(A)⁺ RNA from TE-1. The probes were MB38 for lane 1, MB40 for lane 2, and MB42 for lane 3. (B) RNase protection assay of TE-1. Lane 1, RNase-digested mixture of MB38 and 1 μ g of poly(A)⁺ RNA from TE-1; lane 2, RNase-digested MB38; lane 3, RNase-digested mixture of MB40 and RNA from TE-1; lane 4, RNase-digested MB40; lane 5, RNase-digested mixture of MB38, MB42, and RNA from TE-1; lane 6, RNase-digested mixture of MB38 and MB42. (C) PCR amplification from a cDNA mixture of TE-1. Lane 1, 500 ng of *Msp* I-digested pBR322 DNA; lane 2, PCR products between 38A and 38R; lane 3, PCR products between 40A and 40R.

fication of the *HST1* gene and the *INT2* gene in gastric cancers (28). These two genes, however, are rarely expressed in human cancers. Other genes relevant to human cancers may be found in this amplification unit. Recently, the *PRAD1/cyclin D1/EXP2* gene was isolated (refs. 29 and 30; unpublished observation). It is activated by chromosomal inversion in parathyroid adenomas (31). It is a potential candidate for the responsible gene on the amplification unit. The following points, however, should be taken into consideration before drawing a conclusion on the significance of the *PRAD1* amplification in development of cancers. First, the *PRAD1* gene is not amplified in some cancers, while other DNA markers at chromosome 11q13 do show amplification in other cancers (32, 33). The second is the presence of yet-to-be-characterized transcribed sequences (*EXPI*) that map telomeric of *PRAD1* (unpublished data). Furthermore, by using the exon-trapping system, we have isolated three transcribed sequences from the region between *PRAD1* and *HST1*.

Most transcripts undergo RNA processing, and genomic DNA fragments are prepared by partial digestion with *Sau3AI* in our SETS; accordingly, exons with flanking introns will be inserted into the pMHC2 vector in the sense orientation frequently enough. Sequence analysis proved that every cloned fragment had undergone RNA splicing precisely. As demonstrated in the above results, the SETS attains sufficient sensitivity and specificity to screen genomic DNA segments of several hundred kilobase pairs by a single transfection, which implies that YACs can also be analyzed without further subcloning into cosmid vectors.

If this system is applied to the region where detailed genetic or physical maps have been constructed, the large genomic DNA segments will be effectively and easily screened for transcribed sequences. Possible application of the SETS to the regions saturated with YAC clones may allow reverse genetics to make rapid progress. A breakthrough in reverse genetics would be a great help in identifying and isolating genes responsible for human cancers and human genetic disorders.

We thank Dr. T. Nishihira for providing the TE-1 and TE-10 cell lines. This work was supported in part by a Grant-in-Aid for a Comprehensive 10-Year Strategy for Cancer Control from the Ministry of Health and Welfare; by Grants-in-Aid for Cancer Research from the Ministry of Health and Welfare and from the Ministry of Education, Science and Culture of Japan; and by the Bristol-Myers Squibb Foundation.

- Compton, D. A., Weil, M. M., Jones, C., Riccardi, V. M., Strong, L. C. & Saunders, G. F. (1988) *Cell* **55**, 827–836.
- Rose, E. A., Glaser, T., Jones, C., Smith, C. L., Lewis, W. H., Call, K. M., Minden, M., Champagne, E., Bonetta, L., Yeger, H. & Housman, D. E. (1990) *Cell* **60**, 495–508.
- Call, K. M., Glaser, T., Ito, C. Y., Buckler, A. J., Pelletier, J., Haber, D. A., Rose, E. A., Kral, A., Yeger, H., Lewis, W. H., Jones, C. & Housman, D. E. (1990) *Cell* **60**, 509–520.
- Wallace, M. R., Marchuk, D. A., Anderson, L. B., Letcher, R., Odeh, H. M., Saulino, A. M., Fountain, J. W., Brereton, A., Nicholson, J., Mitchell, A. L., Brownstein, B. H. & Collins, F. S. (1990) *Science* **249**, 181–186.
- Cawthon, R. M., Weiss, R., Xu, G., Viskochil, D., Culver, M., Stevens, J., Robertson, M., Dunn, D., Gesteland, R., O'Connell, P. & White, R. (1990) *Cell* **62**, 193–201.
- Joslyn, G., Carlson, M., Thliveris, A., Albertsen, H., Gelbert, L., Samowitz, W., Groden, J., Stevens, J., Spurio, L., Robertson, M., Sargeant, L., Krapcho, K., Wolff, E., Burt, R., Hughes, J. P., Warrington, J., McPherson, J., Wasmuth, J., Paslier, D. L., Abderrahim, H., Cohen, D., Leppert, M. & White, R. (1991) *Cell* **66**, 601–613.
- Kinzler, K. W., Nilbert, M. C., Su, L.-K., Vogelstein, B., Bryan, T. M., Levy, D. B., Smith, K. J., Preisinger, A. C., Hedge, P., Mckechnie, D., Finnear, R., Markham, A., Groffen, J., Boguski, M. S., Altschul, S. F., Horii, A., Ando, H., Miyoshi, Y., Miki, Y., Nishisho, I. & Nakamura, Y. (1991) *Science* **253**, 661–665.
- Monaco, A. P., Neve, R. L., Colletti-Feener, C., Bertelson, C. J., Kurnit, D. M. & Kunkel, L. M. (1986) *Nature (London)* **323**, 646–670.
- Fearon, E. R., Cho, K. R., Nigro, J. M., Kern, S. E., Simons, J. W., Ruppert, J. M., Hamilton, S. R., Preisinger, A. C., Thomas, G., Kinzler, K. W. & Vogelstein, B. (1990) *Science* **247**, 49–56.
- Weber, F., Villiers, J. & Schaffner, W. (1984) *Cell* **36**, 983–992.
- Allen, N. D., Cran, D. G., Barton, S. C., Hettle, S., Reik, W. & Surani, M. A. (1988) *Nature (London)* **333**, 852–855.
- Gossler, A., Joyner, A. L., Rossant, J. & Skarnes, W. C. (1989) *Science* **244**, 463–465.
- Bird, A. P. (1986) *Nature (London)* **321**, 209–213.
- Liu, P., Legerski, R. & Siciliano, M. J. (1989) *Science* **246**, 813–815.
- Corbo, L., Maley, J. A., Nelson, D. L. & Caskey, C. T. (1990) *Science* **249**, 652–655.
- Duyk, G. M., Kim, S., Myers, R. M. & Cox, D. R. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 8995–8999.
- Buckler, A. J., Chang, D. D., Graw, S. L., Brook, J. D., Harber, D. A., Sharp, P. A. & Housman, D. E. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 4005–4009.
- Dominski, Z. & Kole, R. (1991) *Mol. Cell. Biol.* **11**, 6075–6083.
- Naora, H. & Deacon, N. J. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 6196–6200.
- Sakamoto, H., Mori, M., Taira, M., Yoshida, T., Matsukawa, S., Shimizu, K., Sekiguchi, M., Terada, M. & Sugimura, T. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 3997–4001.
- Nishihira, T., Kasai, M., Mori, S., Watanabe, T., Kuriya, Y., Suda, M., Kitamura, M., Hirayama, K., Akaishi, T. & Sasaki, T. (1979) *Gann* **70**, 575–584.
- Sakamoto, H., Yoshida, T., Nakakuki, M., Odagiri, H., Miyagawa, K., Sugimura, T. & Terada, M. (1988) *Biochem. Biophys. Res. Commun.* **151**, 965–972.
- Maniatis, T., Fritsch, E. F. & Sambrook, J. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY), 2nd Ed.
- Okayama, H. & Berg, P. (1982) *Mol. Cell. Biol.* **2**, 161–170.
- Maxam, A. M. & Gilbert, G. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 560–564.
- Wieringa, B., Meyer, F., Reiser, J. & Weissman, C. (1983) *Nature (London)* **301**, 38–43.
- Dominski, Z. & Ryszard, K. (1992) *Mol. Cell. Biol.* **12**, 2108–2114.
- Yoshida, M. C., Wada, M., Satoh, H., Yoshida, T., Sakamoto, H., Miyagawa, K., Yokota, J., Koda, T., Kakinuma, M., Sugimura, T. & Terada, M. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 4861–4864.
- Motokura, T., Bloom, T., Kim, H. G., Juppner, H., Ruderman, J. V., Kronenberg, H. M. & Arnold, A. (1991) *Nature (London)* **350**, 512–515.
- Matsushime, H., Roussel, M. F., Ashmun, R. A. & Sherr, C. J. (1991) *Cell* **65**, 701–713.
- Rosenberg, C. L., Kim, H. G., Shows, T. B., Kronenberg, H. M. & Arnold, A. (1991) *Oncogene* **6**, 449–453.
- Schuuring, E., Verhoeven, E., Mooi, W. J. & Michalides, R. J. A. M. (1992) *Oncogene* **7**, 355–361.
- Proctor, A. J., Coombs, L. M., Cairns, J. P. & Knowles, M. A. (1991) *Oncogene* **6**, 789–795.
- Reed, R. (1989) *Genes Dev.* **3**, 2113–2123.
- Reed, R. & Maniatis, T. (1988) *Genes Dev.* **2**, 1268–1276.