Proc. Natl. Acad. Sci. USA
Vol. 90, pp. 4364–4368, May 1993
Genetics

An integrated approach for identifying and mapping human genes

(cDNA libraries/yeast artificial chromosomes/homologous recombination/expressed sequence tags/human gene mapping)

RUCHIRA DAS GUPTA*, BERNICE MORROW*, IVONNE MARONDEL*, SATISH PARIMOO†, VITA L. GOEI†‡, JEFFREY GRUEN†‡, SHERMAN WEISSMAN†, ARTHUR SKOULTCHI§, AND RAJU KUCHERLAPATI*¶

Departments of *Molecular Genetics and §Cell Biology, Albert Einstein College of Medicine, Bronx, NY 10461; and †Boyer Center for Molecular Medicine and ‡Department of Pediatrics, Yale University School of Medicine, New Haven, CT 06536

Contributed by Sherman Weissman, February 10, 1993

ABSTRACT **We have developed a method for generating expressed-sequence maps of human chromosomes. The method involves several steps that begin with libraries of highly representative short cDNAs prepared by using random oligomers as primers. The cDNA inserts are amplified by PCR with flanking vector primers. Chromosomal region-specific cDNA packets are prepared by hybridization of the cDNA inserts to DNA derived from yeast artificial chromosomes (YACs) assigned to defined regions of human chromosomes. The cDNA packets are cloned into yeast chromosome fragmentation vectors and used for transformation of yeast bearing the YAC used for affinity purification. Sequences in the cDNAs undergo homologous recombination with the corresponding exons in the genomic DNA yielding a set of truncated YACs. Each unique truncation specifies the location of an exon in the YAC. Since all of the truncation events end with the same vector sequence, it is possible to rescue and sequence these ends to generate expressed sequence tags. The method couples rapid purification of region-specific cDNAs with precise mapping of their genes on YACs. Appropriately truncated YACs also provide easy access to gene regulatory sequences. We describe the feasibility of individual steps of the method using the factor IX (*F9*) gene as a model system and we present the mapping of several expressed sequences corresponding to a 330-kb YAC containing DNA from human chromosome 6p21. In addition, we obtained the sequence, including an intron–exon junction, flanking a particular truncation event.**

The human genome is estimated to encode 100,000 genes (1). The coding regions of the genes constitute only 4–5% of the genome. The global human genome effort is designed ultimately to identify all of the genes within the genome. One approach to identifying all the genes is to sequence the genome in its entirety. Although methods to sequence large segments of DNA are being developed, very few large-scale efforts to sequence the human genome are underway. Several different methods, during the past 20 years or so, have led to identification and mapping of nearly 3000 genes (2). In most of these efforts, the process of gene discovery and gene mapping constitute independent programs. For example, Adams *et al.* (3) have described a method for rapidly sequencing short cDNA fragments with an average of 400 bp from tissue-specific cDNA libraries. The sequence obtained is referred to as an expressed sequence tag (EST). Some of the ESTs were assigned to individual human chromosomes by using the probes to screen a panel of mammalian somatic cell hybrids. Although the generation of ESTs is rapid and very useful, this approach does not permit precise mapping of the gene. We have developed a strategy which provides an integrated approach to identifying and mapping genes in the human genome.

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.

The method we describe begins with the construction of randomly primed short-fragment cDNA libraries in λgt10 from several fetal and adult human tissues. The inserts from such libraries are prepared by the polymerase chain reaction (PCR) and size-fractionated. The cDNAs greater than 300 bp are then hybridized to immobilized DNA of a yeast artificial chromosome (YAC) containing a defined human DNA fragment (4). The cDNAs that hybridize to the YAC DNA after two rounds of selection are amplified by PCR and cloned into a yeast chromosome fragmentation vector (CFV) (5). The cDNAs in the CFV are then used to localize the corresponding gene sequences by homologous recombination in yeast. We describe the feasibility of this method by first testing it with a cloned clotting factor IX (*F9*) cDNA to target individual exons within the *F9* gene on a 650-kb YAC. We have also used the procedure to map several cDNAs that were isolated from spleen and duodenum cDNA libraries by hybridization selection with a 330-kb YAC from the major histocompatibility complex (MHC). By using one of the truncated YACs produced in this series of experiments, we were able to rescue and sequence an exon–intron junction from a novel gene on chromosome 6p21.

MATERIALS AND METHODS

Yeast Strains and Media. Experiments described in this report were conducted with two YACs. One is a 650-kb YAC present in the strain YPH 599 (6). The second is a YAC with a 330-kb insert (B30H3) from the class I region of the human *HLA* locus (7). B30H3 was originally in the AB1380 strain of yeast and was transferred to a his3 *Saccharomyces cerevisiae* strain, YPH 252 (*MATa ura 3-52 lys2-801 ade2-101 trp1Δ1 his3Δ200*), as described (6). The name B30H3 was retained for the *his3Δ200* segregant. YPH 599 and B30H3 were maintained in AHC medium [per liter: 1.7 g of yeast nitrogen base (Difco 0335-15), 5 g of $(NH_4)_2SO_4$, 10 g of casein hydrolysate (United States Biochemical catalogue no. 12852), 20 g of glucose, and 20 mg of adenine sulfate]. Yeast transformation was accomplished by the lithium acetate procedure (8) with 1 μg of linearized plasmid. Transformants were selected on synthetic minimal medium (SD) plates lacking tryptophan and histidine (SD−Trp−His) (9). Colonies that survived this selection were replica plated on SD plates lacking uracil (SD−Ura) to identify those colonies which had lost the ability to grow in the absence of uracil. His^+ Trp^+ Ura^- colonies were maintained in SD−Trp−His medium.

Construction of a Composite cDNA Library. Polyadenylylated RNA was isolated from various tissues and used as a substrate for cDNA synthesis by the use of random oligomers

Abbreviations: CFV, chromosome fragmentation vector; EST, expressed sequence tag; *F9*, factor IX (gene); MHC, major histocompatibility complex; PFGE, pulsed-field gel electrophoresis; YAC, yeast artificial chromosome.

¶To whom reprint requests should be addressed.

as primers for the reverse transcriptase reaction (10). For the experiments described in this report cDNA libraries constructed from adult spleen and adult duodenum were used. Each library initially contained more than two million independent clones.

Selection of a cDNA Packet. The selection of spleen cDNA library with the YAC B30H3 has been described (4). A similar selection was carried out with a short-fragment random-primed duodenum cDNA library. In brief, cDNAs of >300 bp were obtained by size fractionation of PCR-amplified inserts from random-primed cDNA libraries cloned in λgt10 by using the vector outer primer set C (C1; 5′-CCACCTTTTGAGCAAGTTCAG-3′; C2; 5′-GAGGTGGCTTATGAGTATTTC-3′) for PCR amplification (4). Chromosomes from the yeast carrying the appropriate YAC were fractionated by pulsed-field gel electrophoresis (PFGE), agarose containing the YAC was excised after digestion with a restriction endonuclease, and the DNA fragments were electroeluted, denatured, and immobilized on filters. The cDNAs were hybridized with the filters containing YAC DNA under conditions that suppress hybridization of repetitive elements. To reduce nonspecific hybridization, the cDNAs hybridizing to the filter after the first round of selection were amplified with λgt10 inner primer set C, size-fractionated, and then rehybridized with the YAC DNA as before. The cDNAs that were retained on the filter after the second round of hybridization were amplified with primer set A (A1, 5′-AGCCTGGTTAAGTCCAAGCTG-3′; A2; 5′-CTTCCAGGGTAAAAAGCAAAAG-3′) and cloned into λgt10.

Vectors and cDNA Packet Construction. Two different CFVs were used in the experiments. For fragmentations involving the *F9* cDNA, a CFV (pRDht) was constructed by inserting a 270-bp human telomeric repeat (provided by R. Moyzis, Los Alamos National Laboratories) into the *Pst* I site in the polylinker of pRSN303 (6). Subsequently, a partial cDNA for *F9* (11) was cloned into the *Xba* I site of pRDht to yield pRDF9ht. This plasmid was linearized at the *Sma* I site for transformation.

The CFV pBP103 (12) was used for fragmentation involving cDNAs isolated from B30H3. Individual cDNAs were amplified from the λgt10 libraries of twice-selected cDNAs by using primer set A and were digested with *Eco*RI and cloned into the single *Eco*RI site of pBP103. In an effort to ensure that cDNAs were present in both orientations, a mixture of 10 transformants for each cDNA was used to transform B30H3, following digestion with *Sal* I. Transformants were treated as described above.

PFGE and Southern Analysis. DNA for PFGE was prepared by methods previously described (13), and DNA for conventional Southern analysis was prepared as described (14). Electrophoretic karyotypes were examined by contour-clamped homogeneous electric field (CHEF; Bio-Rad CHEF DR II apparatus). Routinely, 20-hr runs were performed with a 60-sec pulse time (YPH 599 and its derivatives) or by ramping from 20 to 40 sec (B30H3 and its derivatives). The size of the YACs was determined by transfer of the yeast DNA to filter membranes (Gene Screen *Plus*, NEN; or Hybond-N, Amersham) for Southern blot hybridization with Cot-1 DNA (BRL) as the probe. Other probes used for hybridization included a 1.4-kb *Eco*RI–*Bam* HI fragment from p288 containing the yeast *HIS3* gene (gift of Philip Hieter, Johns Hopkins University) and a 445-bp *Hae* II fragment from pUC19 containing the *lacZ* gene. Radiolabeled probes were prepared by random primer extension (15).

Rescue of End Sequences from Truncated YACs and Sequencing. "Vectorette" PCR was performed to rescue the DNA from the truncated end of YACs. Genomic DNA from yeast containing the truncated YAC was digested with *Ava* II, an enzyme that cuts proximal to the telomere sequences in pBP103. Two hundred fifty nanograms of digested genomic DNA was ligated overnight at 16°C to 1 μg of annealed vectorette (16) with compatible ends in a final volume of 15 μl. PCR was carried out with a primer corresponding to sequences within the vectorette (16) and a second primer (T7) corresponding to sequences adjacent to the *Eco*RI cloning site in pBP103. Cycling conditions were 30 cycles of denaturation at 94°C for 1 min, annealing at 58°C for 1 min, and extension at 72°C for 2 min, followed by a final extension of 10 min at 72°C. The PCR products were cloned into pBluescript II KS (Stratagene) and sequenced (Sequenase 2.0; United States Biochemical) by conventional methods.

RESULTS

Strategy. The overall strategy to isolate and map genes is shown in Fig. 1*a*. Short-fragment (200–400 bp) cDNA libraries prepared from poly(A)$^+$ mRNA derived from specific tissues have been constructed (10). The inserts in these cDNA libraries, homologous to specific regions of the genome, are purified by hybridization selection with a YAC containing human genomic sequences (4). The purified cDNAs can be sequenced directly to generate ESTs. They can also be cloned into a yeast CFV (5, 12) and used to map the corresponding exons in the YAC. The CFV contains selectable markers for propagation in bacteria (*amp*^{R}) and transformation of yeast (*HIS3*), a cloning site for cDNAs, and sequences capable of functioning as a telomere in yeast. When linearized, the plasmid has the cDNA at one end and a functional telomere at the other end. Upon transformation of the yeast strain containing the YAC with the cDNA containing CFV, it is expected that homologous recombination between the individual exon sequences in the cDNA and their corresponding sequences in the YAC will result in a set of truncated YACs in which the sites of recombination represent the locations of exons (Fig. 1*b*). Since each truncated YAC is terminated by common vector sequences, it is possible to sequence that end of the YAC and obtain gene coding sequences which can constitute ESTs. Examination of the truncated YACs and sequencing from the truncated ends permit detection of exons and exon–intron junctions and permit one to deduce the orientation of the gene within the YAC. We show the feasibility of this approach and present the results of mapping of several cDNAs isolated by hybrid selection with a 330-kb YAC from the MHC region.

Testing the Feasibility of Targeting Individual Exons in a Gene. An important aspect of the proposed strategy is the ability to target individual exons in a gene with the corresponding cDNA cloned into a CFV. One concern was that only the exon adjacent to the site of linearization would initiate recombination. We also wanted to investigate whether recombination would be biased by the size of the exon initiating the homologous recombination. We examined these issues by using a partial *F9* cDNA as a recombination substrate and a 650-kb YAC (HYA32G5) containing the *F9* genomic locus as a target (17). The *F9* gene has been localized to a position about 40 kb from the end of the long arm of the YAC and is oriented in the 5′ → 3′ direction with respect to the centromere and *URA3* marker. The complete nucleotide sequence and exon–intron organization of the *F9* gene are known (18). The gene encompasses 33.5 kb of DNA including eight exons, the largest being exon 8 (1936 bp). An *F9* cDNA fragment that lacks exon 1, the first 24 bp of exon 2, and the last 1795 bp of exon 8 (11) was used for the fragmentation studies. As shown in Fig. 1*b*, recombination between individual exons in the cDNA and the corresponding exons in the gene is expected to yield a series of fragmented YACs, the smallest arising from recombination at the most centromeric exon (here exon 2, since the cDNA used lacks exon 1) and the largest arising from recombination at exon 8, which is the most telomeric.

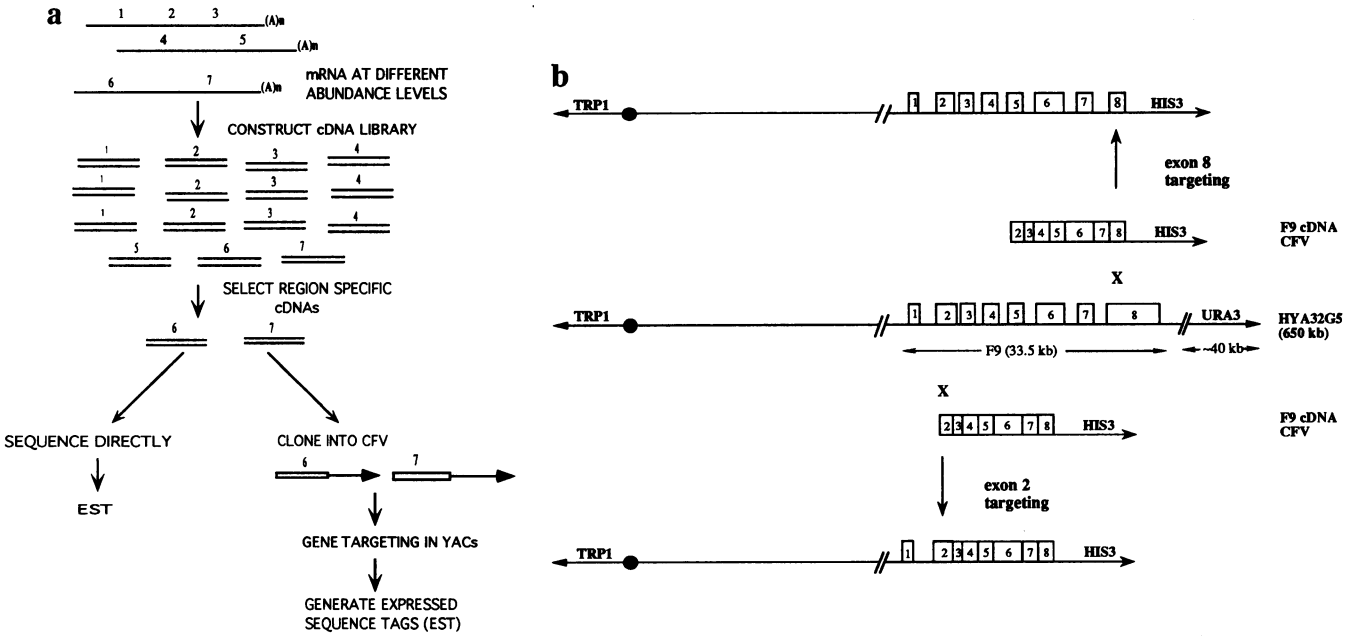

FIG. 1. (*a*) Strategy for isolating and mapping genes. (*b*) Strategy for targeting individual exons within a gene located on a YAC.

The partial *F9* cDNA was cloned into pRDht containing human telomeric sequences, *amp*^{R}, *neo*, and *HIS3* as selectable markers in bacteria, mammalian cells, and yeast, respectively. The his3 yeast strain YPH 599, containing the 650-kb YAC, was transformed with the CFV linearized at the single *Sma* I site. His$^+$ colonies were isolated and examined for loss of the telomeric *URA3* YAC marker by replica plating on SD−Ura plates. Over 90% of the His$^+$ colonies had lost the *URA3* marker, indicating that they had undergone homologous recombination resulting in terminal deletion of the YAC. This was confirmed by PFGE of chromosomes from the yeast carrying the truncated YACs. The size of the individual YACs was determined by Southern blot hybridization of the pulsed-field gel blot with repetitive human DNA (COT-1) as a probe. All of the His$^+$ Ura$^-$ colonies contained YACs smaller in size than the parental YAC. The sizes of these YACs were ≈600 kb, suggesting that they had undergone terminal deletions at or within the *F9* gene, which is 40 kb from the telomere (data not shown).

The large size of the truncated YACs relative to the size of the *F9* gene prevented us from distinguishing among targeting events at different exons of the *F9* gene by PFGE. However, by restriction enzyme mapping and Southern blotting using the known map of the *F9* gene, we were able to identify targeting events within the individual *F9* exons in the YAC. Table 1 shows the sizes of DNA fragments predicted to be present in different targeting events when genomic DNA is digested with *Bgl* II followed by Southern blot hybridization with a *HIS3* probe. Representative results from this type of analysis are shown in Fig. 2. Since the *HIS3* gene in the vector contains *Bgl* II sites, all the transformants have a common band (slightly larger than 4.8 kb and somewhat variable depending on the degree of telomeric repeats) arising from hybridization of the *HIS3* probe to the vector terminated by the telomere. In addition, all transformants have a band of variable size (depending on the exon at which recombination initiated) arising from the vector sequences adjacent to the cloned cDNA. There is also a common 9-kb band which arises from hybridization of the *HIS3* probe to the remaining 25 bp of the chromosomal *his3Δ200* locus (Fig. 2, lane A). The presence of a 1.5-kb band (Fig. 2, lanes B and E) is indicative of targeting to exon 8, a 2.7-kb band (lanes D and F) shows targeting to exon 5, a 3.2-kb band (lane C) is evidence for exon 4 targeting, and a 4.6-kb band (lane G) indicates that targeting initiated at exon 2.

Fifty-nine independent truncation events were analyzed in this fashion. The results, summarized in Table 1, show that six of the seven exons represented in the *F9* cDNA were targeted in the YAC. No targeting events in the smallest of the *F9* exons, exon 3 (25 bp), were detected. The results did not show any bias toward recombination involving exon 2,

Table 1. Summary of *F9* cDNA targeting data

Exon no.	Size, bp	Fragment size, kb	No. of colonies	% total
2	141	4.6	16	27.1
3	25	4.2	0	0.0
4	114	3.2	5	8.5
5	129	2.7	15	25.4
6	204	5.3	7	11.9
7	114	3.1	3	5.1
8	141	1.5	13	22.0

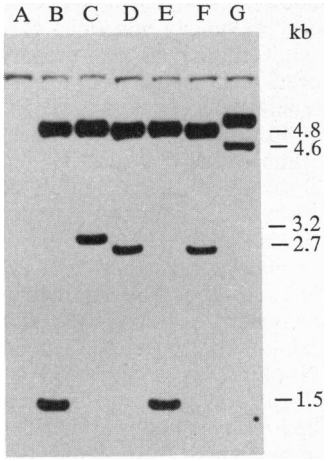

FIG. 2. Detection of targeting events to specific exons in the *F9* gene. *Bgl* II-digested genomic DNA was Southern blotted and hybridized with a *HIS3*-specific probe. Lane A, YPH 599; lanes B–G, independent terminally deleted transformants of YPH 599 with *F9* cDNA-containing CFV.

which is adjacent to the site of linearization of the CFV, since exons 2 and 8 were targeted with roughly equal frequency. All exons in the range 114–204 bp were targeted. These results indicate that it is possible to target individual exons within a gene by using cDNAs cloned into a CFV.

Targeting with B30H3-Selected cDNAs. Having demonstrated the feasibility of targeting a gene on a YAC with a cloned cDNA, we proceeded to use cDNAs isolated by hybridization selection with B30H3 to assign their map positions on the YAC. Six cDNAs—VLG002, VLG003, VLG007, VLG010, VLG052, and VLG058—were isolated from a duodenum cDNA library (V.L.G. and J.G., unpublished work). Hybrid selection of the same YAC with a spleen cDNA library had yielded a cDNA, B30.7, for a novel gene as well as partial cDNAs for the *HLA-A* gene (4). These eight cDNAs were cloned into pBP103 and used individually to transform B30H3 as described. We found that with each cDNA, 75–90% of the His^+ clones were Ura^-, indicating terminal deletions had occurred at a high frequency with each cDNA. Based on the size of the truncated YACs obtained with each cDNA, their sites of recombination can be ordered along the YAC, with VLG007 being the most centromeric, followed by VLG058, VLG010, VLG052, B30.7, VLG003, and VLG002. The sizes of the truncated YACs range from ≈285 kb to ≈55 kb. The *HLA-A* cDNA yielded several classes of truncation events. Since four *HLA-A* pseudogenes are known to be located near the functional gene (7), these results indicate targeting of *HLA-A* cDNA to the normal as well as some of its pseudogenes.

In addition to transformations with individual cDNAs, a reconstruction experiment was carried out by mixing equal quantities of the individual cDNA containing CFVs and using the mixture to transform B30H3. Seventy-five percent of the His^+ colonies derived from transformation were Ura^-, indicating that the efficiency of fragmentation with the cDNA mixture was similar to that observed with individual cDNAs. We observed that all of the size classes of truncated YACs obtained with individual cDNAs were recovered when the cDNA mixture was used (Fig. 3). The results indicate that cDNA packets generated by the affinity-capture methods can be utilized to map their corresponding genomic sequences in YACs.

An individual short fragment cDNA may give rise to only one size class of truncated YACs or to various size classes, presumably depending on the number of exons present in the cDNA and the spacing between the exons. Even when the truncated YACs are of uniform size by PFGE analysis, it is possible that they represent targeting to different closely spaced exons. To ascertain whether this were the case,

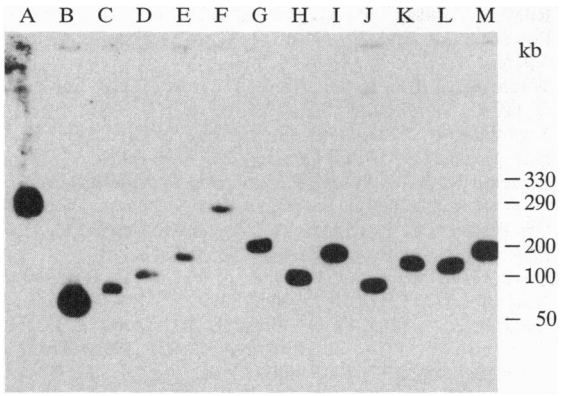

FIG. 3. Truncation events generated by transformation of B30H3 with an artificial mixture of eight different B30H3 YAC-selected cDNAs. Electrophoretic karyograms were obtained by PFGE and the positions of the truncated YACs were determined by Southern blotting and hybridization with a *lacZ* probe.

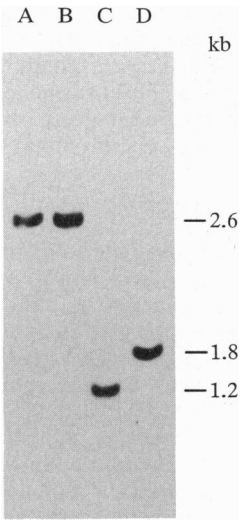

FIG. 4. Detection of different targeting events within members of one size class of truncated YACs. Genomic DNA from four independent transformants derived from targeting of B30H3 with B30.7 cDNA was digested with *Ava* II, Southern blotted, and hybridized with a *lacZ* probe.

genomic DNA from several of the His^+ Trp^+ Ura^- transformants, containing a 150-kb YAC obtained with B30.7 cDNA targeting, were digested with *Hin*dIII and blotted with a *lacZ* probe. Three distinct bands of 2.6 kb, 1.8 kb, and 1.2 kb were obtained, suggesting that recombinations involving different exons of the gene were represented in the truncated YACs (Fig. 4).

To analyze these targeting events in greater detail, we attempted to rescue and sequence DNA near the recombination site in one of the truncated YACs generated with B30.7 cDNA. Genomic DNA from the strain shown in Fig. 4, lane D, was digested with *Ava* II, ligated to a vectorette with compatible ends, and amplified by PCR as described in *Materials and Methods*. A 520-bp PCR product was obtained. Nucleotide sequence of this fragment was obtained and compared with that of B30.7 (Fig. 5). The sequence (251 nucleotides) derived from the terminal end of the truncated YAC was identical to that of the B30.7 cDNA, whereas the rest of the sequence, except for the first 8 nucleotides, was different. These 8 nucleotides are adjacent to an *Ava* II site in a second exon of B30.7. The region of divergent sequence is terminated by splice donor (GT) and acceptor (AG) dinucleotides, and there is a pyrimidine-rich stretch of nucleotides and a branch-point consensus sequence upstream from the putative splice acceptor site indicating that this sequence corresponds to an intron. The position and sequence of the intron were confirmed by isolating and sequencing an incompletely spliced cDNA from an oligo(dT)-primed thymus

GACCTGTGgtaagctaatgagatcaagaactggctccataaggtgggtaggaaagaaatggagga

gtgattgcaaagctctggagagttttgtgcccaattccaagagggaaaagagatgtaaaccatcgacgt

ttgagaggcgtgatcgcctgattcctgtgggaagtaaggggatatgaccaggcctcctctaaccaccag

tttcttcccagGTTGACAGGCGCTGCCCTCGATGTGGTCATGAAGGAATGGCATACC

ACACCAGACAGATGCGTTCAGCCGATGAAGGGCAAACTGTCTTCTACACCTGT

ACCA ACTGCAAGTTCCAGGAGAAGGAAGACTCTTGACCTTTTTCCTGGGCAAC

TCTACAGTCCCTCCCTCCTTTCGGAAGGTGCAGGATACTGGGTTTTTAGATGCCT

TGTCCATCCTGTCTGGTTGCAATGTTTTGCTCCCAG

FIG. 5. Sequence of the DNA rescued from a B30.7 truncation event. Exon sequences are shown in uppercase letters and intron sequences in lowercase letters. AG/GT splicing and branch-point consensus signal sequences are underlined.

cDNA libraries (unpublished results). These results clearly show that an exon in the B30.7 cDNA underwent homologous recombination with its corresponding exon in the B30H3 YAC. The recovery of a consensus splice acceptor site also allows us to establish the orientation of the gene on the YAC, placing the 5′ end of the gene proximal to the centromere.

DISCUSSION

We have developed and evaluated a method to isolate cDNA sequences corresponding to defined segments of human DNA cloned into a YAC and to map their relative positions along the YAC. The mapping strategy involves cloning the cDNA fragments into a yeast CFV and using the resulting plasmids to transform yeast containing the YAC used for the purification of the cDNAs. We evaluated the mapping strategy first by determining whether individual exons within the *F9* gene on a 650-kb YAC could be targeted by their corresponding sequences in a *F9* cDNA cloned in a CFV. The *F9* cDNA clone used for targeting encompassed seven exons ranging in size from 25 bp (exon 3) to 214 bp (exon 6). Examination of 59 targeted events showed that all of the exons except the smallest, the 25-bp exon 3, were targeted. Since homologous recombination with 25 bp of homology or less has been reported in yeast (19), the failure to detect targeting events in exon 3 very likely reflects a relatively low rate of recombination with 25 bp of homology and the limited number of transformants examined.

It is known that double-strand breaks in the region of homology enhance homologous recombination in yeast (20). Therefore, it was possible that the majority of recombination events would initiate within exon 2, near the sequences immediately proximal to the site of linearization in the vector. However, we found that exon 5, which is about the same size as exon 2 but located internally in the cDNA, targeted with about equal efficiency, as did exon 8, which is even further from the site of linearization. Whatever the mechanisms of recombination, our results indicate that the location of the exon within the cDNA does not have any significant effect on the frequency of homologous recombination.

We have shown that several individual cDNA fragments can be targeted to their corresponding genes in a 330-kb YAC. Reconstruction experiments with a mixture of eight different cDNA fragments revealed that all of the different expected size classes of truncation events were recovered. These results indicate that uncharacterized, affinity-purified cDNA packets corresponding to a YAC can be used to deduce the location of their corresponding genes. Such an approach might permit rapid identification of genes encompassed by a YAC.

The method we have described permits different levels of detailed analysis of the cDNA targeting events. Examination of the recombination products by PFGE reveals clusters of exons, and restriction enzyme digestion followed by blot hybridization of individual members within a PFGE size class might reveal targeting events to individual exons within a cluster. If a complete cDNA is used, this method will permit an accurate measure of the number of exons within a gene. Even greater resolution can be achieved by rescuing the ends of truncated YACs and sequencing them. We obtained the exon–intron junction within a gene corresponding to B30.7 cDNA. The derived sequence diverged from the known sequence of B30.7 at a site that precisely corresponded to an exon–intron junction, and the sequence beyond this junction was confirmed to match the sequence of an intron of B30.7. This same strategy also permits access to sequences at the 5′ or 3′ end of genes (depending on the orientation of the gene on the YAC).

Several methods to identify transcriptional units in fragments of DNA have been described. These include exon trapping (21), exon amplification (22), conserved sequence recognition (23), and search for genes in the vicinity of "*Hpa* tiny fragment" islands (24). The procedure described here provides an additional means of identifying, orienting, mapping, and isolating genes from specific cloned DNA fragments.

The vector pBP103 was kindly provided by Drs. W. Pavan and R. Reeves (Johns Hopkins University), and the human telomeric repeat sequences were provided by Dr. R. Moyzis (Los Alamos National Laboratories). This work was supported by grants from the National Institutes of Health (GM33945 and HG00670 to R.K., CA16368 to A.S., CA42556-07 to S.W., DK34989 and HD27757 to J.G., and HD07388 to V.L.G.), the Wills Foundation (R.K.), the Charles H. Hood Foundation (J.G.), and the Swebilius Cancer Research Fund (J.G.) and by Cancer Center Grant CA13330 to the Albert Einstein College of Medicine.

1. Bodmer, W. F. (1989) *Cold Spring Harbor Symp. Quant. Biol.* **51**, 1–6.
2. Human Gene Mapping 11 (1991) *Cytogenet. Cell Genet.* **58**.
3. Adams, M., Kelley, J. M., Gocayne, J. D., Dubnick, M., Polymeropoulos, M. H., Xiao, H., Merril, C. R., Wu, A., Olde, B., Moreno, R. F., Kerlavage, A. R., McCombie, W. R. & Venter, J. C. (1991) *Science* **252**, 1651–1656.
4. Parimoo, S., Patanjali, S. R., Shukla, H., Chaplin, D. D. & Weissman, S. M. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 9623–9627.
5. Vollrath, D., Davis, R. W., Connelly, C. & Hieter, P. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 6027–6031.
6. Campbell, C., Gulati, R., Nandi, A. K., Floy, K., Hieter, P. & Kucherlapati, R. S. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 5744–5748.
7. Geraghty, D. E., Pei, J., Lipsky, B., Hansen, J. A., Taillon-Miller, P., Bronson, S. K. & Chaplin, D. D. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 2669–2673.
8. Ito, J., Fukuda, Y., Murata, K. & Kimura, A. (1983) *J. Bacteriol.* **153**, 163–168.
9. Rose, M., Winston, F. & Hieter, P. (1990) *Methods in Yeast Genetics* (Cold Spring Harbor Lab., Plainview, NY).
10. Patanjali, S. R., Parimoo, S. & Weissman, S. M. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 1943–1947.
11. Jagadeeswaran, P., Lavelle, D. E., Kaul, R., Mohandas, T. & Warren, S. T. (1984) *Somat. Cell Mol. Genet.* **10**, 465–473.
12. Pavan, W. J., Hieter, P., Sears, D., Burkoff, M. & Reeves, R. (1991) *Gene* **106**, 125–127.
13. Schwartz, D. C. & Cantor, C. R. (1984) *Cell* **37**, 67–75.
14. Davis, R. W., Thomas, M., Cameron, J. R., St. John, T. P., Scherer, S. & Padgett, R. A. (1980) *Methods Enzymol.* **65**, 404–411.
15. Feinberg, A. P. & Vogelstein, B. (1983) *Anal. Biochem.* **132**, 6–13.
16. Riley, J., Butler, R., Ogilvie, D., Finniear, R., Jenner, D., Powell, S., Anand, R., Smith, J. C. & Markham, A. F. (1990) *Nucleic Acids Res.* **18**, 2887–2890.
17. Wada, M., Little, R. D., Abidi, F., Porta, G. & Schlessinger, D. (1990) *Genomics* **7**, 363–370.
18. Yoshitake, S., Schach, B. G., Foster, D. C., Davie, E. W. & Kurachi, K. (1985) *Biochemistry* **24**, 3736–3750.
19. Schiestl, R. H. & Petes, T. D. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 7585–7589.
20. Orr-Weaver, T. L., Szostak, J. W. & Rothstein, R. J. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 6354–6358.
21. Duyk, G. M., Kim, S., Myers, R. M. & Cox, D. R. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 8995–8999.
22. Buckler, A., Chang, D. D., Graw, S. L., Brook, J. D., Haber, D. A., Sharp, P. A. & Housman, D. E. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 4005–4009.
23. Monaco, A. P., Nere, R. L., Colletti-Feener, C., Bertelson, C. J., Kurnit, D. M. & Kunkel, L. M. (1986) *Nature (London)* **316**, 336–338.
24. Bird, A. (1987) *Trends Genet.* **3**, 342–347.