SHORT COMMUNICATION

Expressed Sequence Tags and Chromosomal Localization of cDNA Clones from a Subtracted Retinal Pigment Epithelium Library

LINN GIESER* AND ANAND SWAROOP*,†,[1]

*Departments of Ophthalmology and †Human Genetics, Kellogg Eye Center, University of Michigan, Ann Arbor, Michigan 48105

Received February 5, 1992; revised April 6, 1992

Expressed sequence tags (ESTs) provide useful molecular landmarks for physical mapping and identify the position of an expressed region in the genome. The use of subtracted cDNA libraries enriched for tissue-specific genes as a source of ESTs should reduce the repetitive isolation of constitutively expressed sequences. We report here the sequence tags from the 3′-end region of 58 new directionally cloned cDNAs from a subtracted human retinal pigment epithelium (RPE) cell line library. Eight of the cDNAs have been assigned to human chromosomes using PCR-based EST assays. Chromosomal mapping of subtracted RPE cDNA clones may also help in identifying candidate genes for inherited eye diseases.

Short stretches of unique DNA sequences (sequence-tagged sites, STSs) define landmarks on the genome and can be used effectively for physical mapping strategies (5). STS mapping is now recognized as a prerequisite for generating large contig maps (3). In addition to providing the same information as genomic clones, the tagged sequence sites from cDNA clones (expressed sequence tags, ESTs) also mark the position of an expressed gene. The availability of unbiased and high-fidelity directional cDNA libraries with defined orientation should maximize the output of ESTs. The sequence from the 3′-region of a cDNA frequently has no introns and shows significantly higher polymorphism between species and within different members of a family. Therefore, PCR primers from this region are useful for chromosomal localization (12), for developing polymorphic markers using single-strand conformation polymorphism (SSCP) analysis (6) or denaturing gradient gel electrophoresis (4), and for determining the expression pattern of tagged cDNA. The sequence near the poly(A) can also serve as a reference point for comparing various ESTs. Adams *et al.* (1) isolated 348 random ESTs from a brain library. However, this strategy is only useful in the early stages of constructing the expression map of the human genome because of the repetitive isolation of highly represented and constitutively expressed genes for housekeeping functions and because of the developmental and tissue-specific expression of many genes. To reduce the redundant efforts and to obtain a complete expression map, the cDNA libraries from human tissues may be enriched for specific genes by subtraction cloning (2, 8, 11). We have generated subtracted human retina and retinal pigment epithelium (RPE) cell line cDNA libraries by eliminating most of the constitutively expressed genes using a biotin-based procedure (11). Identification of ESTs for 58 novel cDNAs from the subtracted RPE library and chromosomal localization of 8 of these clones using polymerase chain reaction (PCR)-based EST assays are reported here.

A versatile phage vector, Charon BS, has earlier been described for the preparation of directional cDNA libraries and their efficient transfer to Bluescript plasmid (10). Because of a simple linker-ligation strategy, most cDNA clones in Charon BS libraries contain a stretch of dAs at their 3′-end (generally from the poly(A) tail). Directional and size-selected cDNA libraries for more than 10 human tissues and cell lines have been generated in this vector (A. Swaroop, unpublished data). The fidelity of these libraries was evaluated by isolating cDNA clones for both known and novel genes. Using a biotin-based procedure, a human RPE cell line (0041) cDNA library was enriched for specific clones by subtraction against a mixture of RNA from fetal eye and JY lymphoblastoid cDNA libraries (11). The expression analysis of over 25 random cDNA clones from the subtracted library revealed that more than 80% of these are not detected in the JY library, and at least 4 newly identified clones are expressed only in RPE (N. Agarwal, L. Gieser, M. DelMonte, and A. Swaroop, manuscript in preparation).

To generate ESTs from the 3′-ends, random cDNA clones from the subtracted RPE library (in Bluescript plasmid) were sequenced with T3 primer using the dideoxy chain-termination method, as described (9). The comparison of 108 cDNA sequences to GenBank (Version 69.0) and EMBL (Version 28.0) databases using FASTA (7) showed no cDNA clone with entirely repetitive sequence, polyadenylate insert, or mitochondria-

[1] To whom correspondence should be addressed at 540 Kellogg Eye Center, University of Michigan, 1000 Wall Street, Ann Arbor, MI 48105.

Sequence data reported in this article have been deposited with the EMBL/GenBank Data Libraries under Accession Nos. M91216–1222, M91387–1403, M91480–1511, and M92283–2284.

GENOMICS 13, 873–876 (1992)
0888-7543/92 $5.00

encoded sequence. Of these cDNAs, 36 represent 28 unique clones that are identical to human genes or are the human homologs of vertebrate genes (Table 1). Eleven cDNAs are partially homologous to human repeat sequences or to ribosomal RNAs. Among the 61 clones with no homology to databases, only 3 were observed twice. The sequences of 58 new cDNA clones have been submitted to the GenBank Data Library.

The sequence information from eight of the new cDNAs was used to design the PCR primers using the OLIGO primer analysis software (National Biosciences, Plymouth, MN). The oligonucleotides, synthesized on a DNA synthesizer (Applied Biosystems), were used without purification. The PCR amplification was performed using the Gene-Amp kit and a thermocycler from Perkin–Elmer/Cetus. Annealing conditions (shown in Table 2) yielding a human-specific product were optimized before somatic cell hybrid analysis. The amplification products were visualized by ethidium bromide staining after agarose gel electrophoresis (NuSieve 3:1, FMC Bioproducts, Rockland, ME). DNAs from somatic cell hybrid panels (1 and 2) were purchased from NIGMS repository (Camden, NJ). The description of the hybrids is provided in the catalog of the repository. Each of the EST primer pairs was used to amplify a human-specific product from the DNAs of a panel of 18 human–rodent

TABLE 1

Similarity of ESTs to Genes in GenBank/EMBL Databases

Clone name	Identification	Organism	Accession No.	Redundancy	Chromosomal location
AA2, LG29	Cystatin C (CST3)	Human	X52255, M27891	2	20
AA13	Cytochrome c-1	Human	J04444, X06994	1	8
AA19	Centrosomin A	Mouse	X17373	1	
AA22	MIC2, E2 antigen	Human	M16279, X16996	1	Xp22.32, Yp11.3
AA25	C-JUN	Human	J04111	1	1p31–p32
AA27	R-ras	Human	M14949	1	19q13.3–qter
AA34, AA49	Profilin	Human	J03191	2	1, 17p13.3
AA47	Vimentin	Human	M25246, X56134	1	10p13
LG7	Δ^3-Δ^2-Enoyl-CoA isomerase	Rat	M61112	1	
LG17	Rat sperm membrane protein (YWK-2)	Rat	M31322		
LG20, LG95, LG117	80–87 kD Myristoylated alanine-rich C kinase substrate (MARCKS) protein	Bovine	M24638, M23738	3	
LG34	Proliferating cell nuclear antigen (PCNA), Cyclin	Human	J04718, M15796	1	20pter–q13
LG48, LG87	DNA binding protein B (dpb B), Y box binding protein-1 (YB-1)	Human	M24070, J03827	2	
LG49	Pre-mRNA splicing factor SF2p32	Human	M69039	1	
LG68	Glyceraldehyde 3-phosphate dehydrogenase	Human	X01111, J04038	1	
LG75, LG100	Apoferritin H	Human	M11146, X03486, X03488	2	
LG81	Prothymosin-α (ProT-α)	Human	M14483, M26708	1	
LG82, LG168	Apolipoprotein E	Human	M10065, M12529	2	19q13.2
LG101	Unidentified liver protein	Pig	M29072	1	
LG110, LG152	LL Rep 3	Human	X17206	2	
LG135	Nonhistone chromosomal protein HMG-14	Human	M21339, J02621	1	21
LG137	Platelet glycoprotein Ib β chain	Human	J03259	1	
LG151	β-Galactoside-binding lectin	Human	X14829	1	
LG160	3-Alkyladenine DNA glycosylase (HAAG)	Human	M74905	1	16
LG177	jun-D (identity over a 97-bp overlap)	Human	X56681, X51346	1	19p13.2
LG185	Clathrin light chain α	Human	M20471, M20472	1	
LG192	Acidic ribosomal phosphoprotein, P0	Human	M17885	1	
LG194	Ribosomal protein L7A large subunit, PLA-X gene	Human	M36072, X06705	1	9q33–q34

TABLE 2

PCR Primers, Conditions for Amplification, and Chromosomal Localization of Expressed Sequence Tags

Clone name	D number (GenBank No.)	Primer sequence	Product size (bp)	Annealing temperature (°C)	dNTPs (μM)	Chromosomal location
AA12	D11S869E (M91217)	A: AGT GAT CTC TGA GAA AAG GG B: ATA TTG TGA ATG ACT AGG G	277	57	80	11
AA20	D10S181E (M91221)	A: CAG TGT CAG ACA CGT TAT ATT TG B: AGC TTT CTC TCA GTG TTC TGG C	271	55	80	10
AA26	D19S205E (M91222)	A: ATC CGG ATG TCC AGA GCC ACA G B: CGA GAA GCA TTC CAC TTT GTA GC	206	57	80	19
AA29	D20S76E (M91392)	A: CAA AAA CAG GAA GCA GGC ACG B: GCT GTC ACC ACT TGG TCA GAA	215	60	200	20
AA32[a]	D7S469E (M91387)	C: AGT TAC TTT AAC CTG CAC GGG TTC CG E: TCG GGC ATC CTG TGT CGC TG	201	60	200	7
AA35	D3S1231E (M91388)	A: GAC TTT ATT TAG TTC CTT TGG C: TAC AGA TGC AGC CAC AGA TTG	196	50	200	3
AA36	D1S174E (M91389)	A: ACT CCC TTT CCT CTT GAG GCT G B: GAC AGG ACC ATA CCC TCT TTG	223	58	80	1
AA38[a]	D21S249E (M91390)	A: TAG GTG TCG GGA CTG CTC TGC B: CAG CTG GTT CGG GCT GTG ACG	≈800	57	200	21

Note. $MgCl_2$ concentration 1 mM, except 2 mM for AA12 and 3 mM for AA38. 50-μl reactions, Amplitaq 1.25 U/rxn, primer concentration 0.2 μM, DNA template concentration 100 ng/rxn.

[a] 200 ng/rxn template DNA.

somatic cell hybrids (panel 1). Monochromosomic human–rodent hybrid panel 2 was used only to confirm the localization of AA38 cDNA. The presence or absence of the human-specific PCR product was then correlated with the content of human chromosomes in that particular hybrid, and an EST showing no discordancy was then assigned to a particular chromosome. The results of the chromosomal mapping studies and the D-number assignments for the eight new cDNAs are also shown in Table 2.

The major problems in developing a complete EST map of the human genome are the poor quality cDNA libraries and repetitive isolation of constitutively expressed genes. High quality cDNA libraries enriched for tissue-specific genes by subtraction cloning can minimize the redundant efforts and should provide an excellent source for generating new EST markers. The sequence tags for 58 new cDNA clones from a subtracted RPE library are reported here. The sequence analysis confirms that most of the constitutive genes have been eliminated by subtraction. Several of the known genes listed in Table 1 appear to be involved in cell growth and proliferation and are probably expressed in the RPE cell line but not in the tissue *in vivo*.

The chromosomal localization of ESTs using PCR assay provides valuable information for physical and genetic mapping studies. Eight of the new cDNA clones (ESTs) that are mapped in this study can now be used for sublocalization on a particular chromosome. We were unable to assign three additional ESTs to human chromosomes because of a similar size amplification product from the rodent DNA in the human–rodent somatic cell hybrids. Isolation of new ESTs from the subtracted RPE and retina libraries and their chromosomal localization should provide markers for physical mapping and may facilitate the identification of candidate genes for eye diseases.

ACKNOWLEDGMENTS

We thank Dr. Sherman M. Weissman for encouragement and Ms. Diane Sills for technical assistance during the early stages of this project. Thanks are also due to Drs. Neeraj Agarwal, Denise Yan, and Ricardo Fujita for helpful comments, Mr. Shayne Boucher for invaluable computer support, and Ms. Dorothy Giebel and Ms. Lauri Allis for the preparation of the manuscript. This work is supported in part by grants from the National Institutes of Health EY-07961, the George Gund Foundation, and the Retinitis Pigmentosa Foundation, Baltimore (to A.S.).

REFERENCES

1. Adams, M. D., 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., and Venter, J. C. (1991). Complementary DNA sequencing: Expressed sequence tags and human genome project. *Science* **252:** 1651–1656.
2. Duguid, J. R., Rohwer, R. G., and Seed, B. (1988). Isolation of cDNAs of scrapie-modulated RNAs by subtractive hybridization of a cDNA library. *Proc. Natl. Acad. Sci. USA* **85:** 5738–5742.
3. Green, E. D., Mohr, R. M., Idol, J. R., Jones, M., Buckingham, J. M., Deaven, L. L., Moyzis, R. K., and Olson, M. (1991). Systematic generation of sequence-tagged sites for physical mapping of human chromosomes: application of the mapping to human chromosome 7 using yeast artifical chromosomes. *Genomics* **11:** 548–564.
4. Myers, R. M., Lumelsky, N., Lerman, L. S., and Maniatis, T. (1985). Detection of single base substitutions in total genomic DNA. *Nature* **313:** 495–498.

5. Olson, M., Hood, L., Cantor, C., and Botstein, D. (1989). A common language for physical mapping of the human genome. *Science* **245:** 1434–1435.

6. Orita, M., Suzuki, Y., Sekiya, T., and Hayashi, K. (1989). Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. *Genomics* **5:** 874–879.

7. Pearson, W. R. (1990). Rapid and sensitive sequence comparison with FASTP and FASTA. *In* "Methods in Enzymology" (R. F. Doolittle, Ed.), Vol. 183, pp. 63–98, Academic Press, San Diego.

8. Sive, H. L., and St. John, T. (1988). A simple subtractive hybridization technique employing photoactivatable biotin and phenol extraction. *Nucleic Acids Res.* **16:** 10937.

9. Swaroop, A., Hogan, B. L. M., and Francke, U. (1988). Molecular analysis of the cDNA for human SPARC/Osteonectin/BM-40: Sequence, expression, and localization of the gene to chromosome 5q31–q33. *Genomics* **2:** 37–47.

10. Swaroop, A., and Weissman, S. M. (1988). Charon BS (+) and (−), versatile λ phage vectors for constructing directional cDNA libraries and their efficient transfer to plasmids. *Nucleic Acids Res.* **16:** 8739.

11. Swaroop, A., Xu, J., Agarwal, N., and Weissman, S. M. (1991). A simple and efficient cDNA library subtraction procedure: isolation of human retina-specific cDNA clones. *Nucleic Acids Res.* **19:** 1954.

12. Wilcox, A. S., Khan, A. S., Hopkins, J. A., and Sikela, J. M. (1991). Use of 3′ untranslated sequences of human cDNAs for rapid chromosome assignment and conversion to STSs: Implications for an expression map of the genome. *Nucleic Acids Res.* **19:** 1837–1843.