Am. J. Hum. Genet. 49:581–589, 1991

Isolation of Large Numbers of Chromosome 3–specific Cosmids Containing Clusters of Rare Restriction-Endonuclease Sites

William A. Golembieski,* Scott E. Smith,† Franco Recchia,† Andleeb Judge,† Viji Shridhar,† Orlando J. Miller,† Harry Drabkin,‡ and David I. Smith†

*Center for Molecular Biology and †Department of Molecular Biology and Genetics, Wayne State University School of Medicine, Detroit; and ‡University of Colorado Health Sciences Center, Denver

Summary

We tested 519 chromosome 3–specific cosmids for the presence of rare restriction-endonuclease sites in a search for cosmids containing HTF islands. We have identified 49 cosmids (9% of those tested) that contain multiple rare restriction-endonuclease sites. The cosmids were digested with several common cutting restriction endonucleases to liberate small fragments which were tested as unique-sequence chromosome 3–specific hybridization probes and for evolutionary sequence conservation. Unique-sequence hybridization probes isolated from the cosmids were hybridized to a somatic cell hybrid deletion mapping panel to subchromosomally localize the cosmids. Fragments from many of these cosmids demonstrated conservation of sequence through evolution, and these fragments hybridize to distinct transcripts. These cosmids should therefore prove a useful resource for the identification of many chromosome 3–specific genes, in addition to having potential use as linking clones for pulsed-field gel mapping studies.

Introduction

Many known vertebrate genes have distinctive sequences (i.e., HTF islands) surrounding their transcription start sites (Bird 1986). HTF islands contain a high density of CpG and can be detected within chromosomal as well as cloned DNA as clustered sites for certain rare-cutting restriction enzymes (Bird et al. 1985; Brown and Bird 1986). Lindsay and Bird (1987) tested 32 X-specific cosmids for the presence of the relatively inexpensive rare-cutting restriction endonuclease *Sac*II. They identified four cosmids with *Sac*II sites, and these sites were all located in HTF islands. Hybridization to northern blots provided preliminary evidence that three of the islands were associated with genes.

Large numbers of cosmids can thus be easily screened to identify HTF islands, which should help to pinpoint genic sequences. The use of somatic cell hybrids containing defined regions of interest enables one to quickly localize candidate genes in those regions. Recombinants containing HTF islands are also extremely powerful pulsed-field gel mapping tools. Since the rare restriction sites present within the islands are characteristically unmethylated, recombinants containing HTF islands can be used as linking clones in the construction of macrorestriction maps. Brown and Bird (1986) used this technique to construct maps for a total of 1,500 kb of mouse genomic DNA surrounding three distinct islands.

Our laboratory has isolated over 3,000 chromosome 3–specific cosmids and prepared DNA from each of these (Smith et al. 1989*b*). We previously described our analysis of the frequency and distribution of rare restriction endonuclease sites within several megabases of cloned chromosome 3 DNA (Smith et al. 1987). An analysis of 5,360 kb of DNA revealed five cosmids containing clusters of rare restriction sites. We have extended this analysis by testing 519 cosmids (corresponding to 19,000 kb, or approximately 10% of chromosome 3 DNA) to identify 49 cosmids containing HTF islands.

Received March 15, 1991; revision received April 29, 1991.

Address for correspondence and reprints: Dr. David I. Smith, Department of Molecular Biology and Genetics, Wayne State University School of Medicine, 540 East Canfield, Detroit, MI 48201.

0002-9297/91/4903-0011$02.00

In the present report we describe the characterization of these cosmids. Unique-sequence hybridization probes were isolated from all these cosmids and were hybridized to a somatic cell hybrid deletion mapping panel capable of resolving chromosome 3 into eight distinct subregions. Most of these cosmids contain sequences that are evolutionarily conserved, and these sequences hybridize to distinct message transcripts. These cosmids should therefore be a valuable resource for the identification of chromosome 3–specific genes.

Material and Methods

Hybrid Cell Lines

The Chinese hamster ovary (CHO) parent cell line UrdC$^-$ is an auxotrophic mutant which, for growth, requires exogenous uridine or complementation with the long arm of chromosome 3 (Patterson et al. 1983). This cell line was used to generate a group of hybrids which constitute the somatic cell hybrid deletion mapping panel used in the present study (Drabkin et al. 1987, 1990). Figure 1 shows both the portion of chromosome 3 present in each of the hybrids and the resolving capability of the entire deletion mapping panel. All hybrids except for Hy86D13 were grown in F-12 media and 10% dialyzed FCS to select for retention of chromosome 3. Hy86D13 was grown in DMEM and 10% regular FCS.

Restriction-Endonuclease Digestion of the DNA

All restriction enzymes used in the present study were obtained from Bethesda Research Laboratories (BRL). DNA was digested in enzyme buffers specific for each restriction endonuclease as described by BRL, and we generally used four units restriction endonuclease/μg input DNA.

Testing Fragments as Unique Sequence Chromosome 3–Specific Hybridization Probes

Restriction fragments were cut out of LMP-agarose (BRL-Gibco) and then radiolabeled with [^{32}P]-dCTP (3,000 Ci/mmol) by using random hexanucleotide primers (Feinberg and Vogelstein 1983). The radiolabeled fragments were not purified from the random

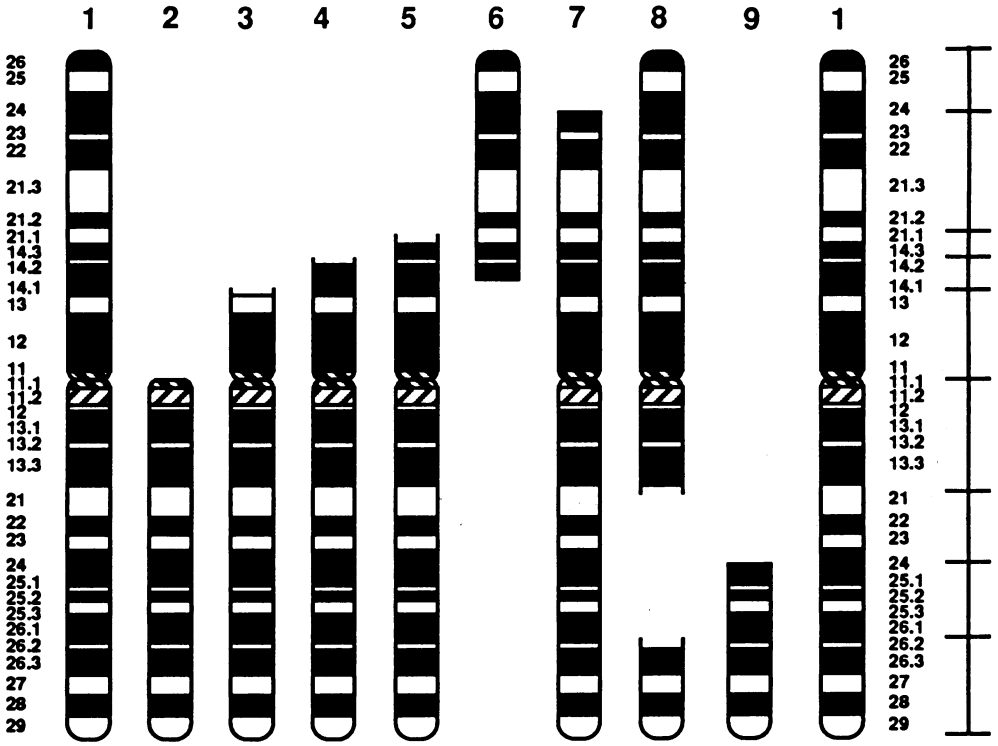

Figure 1 Hybrids constituting somatic cell hybrid deletion mapping panel: 1, UCTP-2A-3 (intact 3); 2, UCTP-2A-3 #5 (3qter-cen); 3, 182UC/7-3 (3p13-qter); 4, TL12-8 (3p14.2-qter); 5, 3;7/UC-1 (distal 3p21.1-qter); 6, R1-1 (3p14.1-pter); 7, 3;21/UC-1 (3p24.2-qter); 8, H3-4 (3pter-q21::q26-qter); and 9, Hy86D13 (3q24.2-qter).

primer–labeling reaction but instead were directly hybridized to Nytran filters (Schleicher & Schuell) that contained *Eco*RI-digested and electrophoretically resolved genomic DNA from (1) CHO, (2) HeLa, and (3) hybrid HHW 423 (whose only human component is an intact chromosome 3). The probes were hybridized to the Nytran filters as described elsewhere (Carlock et al. 1986). Probes that hybridized to HeLa and to HHW 423 DNAs with minimal or no background hybridization were then hybridized to a more extensive panel of DNA—CHO, HeLa, HHW 423, UCTP-2A-3, and Hy137 RT9 (these last three hybrids contain as their only human component a single independently isolated chromosome 3)—to verify that they were chromosome 3–specific hybridization probes.

Hybridizing Unique Sequence Hybridization Probes to the Somatic Cell Hybrid Deletion Mapping Panel

High-molecular-weight DNA was isolated from the various hybrids by using published procedures (Mears et al. 1978). Unique-sequence hybridization probes isolated from the cosmids were first hybridized to Nytran filters containing (1) CHO, (2) HeLa, (3) UCTP-2A-3 #5 (3q-only hybrid), and (4) Hy137RT9 as described elsewhere (Carlock et al. 1986). This hybridization determined whether a probe was derived from the long or the short arm. The chromosome 3 content of the various hybrids constituting the somatic cell hybrid deletion mapping panel is shown in figure 1. Long-arm probes were hybridized to filters containing (1) CHO, (2) HeLa, (3) H3-4 (3pter-q21::q26-qter), (4) Hy86D13 (3q24.2-qter), (5) 2A-3 #5, and (6) Hy137RT9. Short-arm probes were hybridized to filters containing (1) CHO, (2) HeLa, (3) UCTP-2A-3 #5 (3cen-qter), (4) 182UC/7-3 (3p13-qter), (5) TL12-8 (3p14.2-qter), (6) 3;7/UC-1 (distal 3p21.1-qter), (7) R1-1 (3p14.1-pter), (8) 3;21/UC-1 (3p24.2-qter), and (9) Hy137RT9 (intact 3).

Testing Fragments for Sequence Conservation

Fragments were tested for sequence conservation by hybridizing to DNA prepared from UCTP-2A-3, CHO, and RAG (mouse). DNAs were prepared using published procedures (Mears et al. 1978). Prehybridization and hybridization conditions were identical to those used to detect unique-sequence hybridization probes, but we altered the washing conditions to take into account greater mismatches between probes and genomic sequences. Filters were washed four times, for 15 min each, with 2 × SSC, 0.1% SDS at room temperature; twice, for 15 min each, with 1.5 × SSC, 0.1% SDS at 58°C; and then twice, for 15 min each, with 1 × SSC, 0.1% SDS at 58°C.

Northern Analysis

Total RNA was isolated from HeLa cells grown in suspension by the method described by Favaloro et al. (1980). Poly A^+ RNA and poly A^- mRNA was isolated by passage through oligo dT-cellulose (type T3; Collaborative Research). Five micrograms of poly A^+ and poly A^- mRNA was fractionated on glyoxal gels (McMaster and Carmichael 1977) and then transferred to Hybond membranes. Several of the fragments which showed conservation of sequence through evolution were radiolabeled using random hexanucleotide primers and hybridized to these membranes by using conditions identical to those described for testing DNA fragments isolated from these cosmids as unique-sequence hybridization probes.

Results

Testing Cosmids for Rare Restriction-Endonuclease Sites

We tested 519 random chromosome 3–specific cosmids for the presence of *Not*I, *Sfi*I, and *Sst*II (an isoschizomer of *Sac*II) sites. The cosmids were digested with *Eco*RI and were ethanol precipitated, and then a small aliquot was electrophoresed through 0.8% agarose gels to confirm that the digestion was complete. Samples of *Eco*RI-digested cosmids were then digested with *Not*I, *Sfi*I, and *Sst*II, and the samples were run alongside the *Eco*RI digests. Cosmids which contained at least two rare sites were subsequently tested for *Bss*HII, *Mlu*I, *Pvu*I, and *Sal*I sites. We found a total of 49 cosmids which contained multiple rare restriction-endonuclease sites (table 1). The rare restriction sites in the cosmids were found to be clustered within the cosmids, as many of the rare sites were present on the same *Eco*RI fragment.

Isolation of Unique Sequence Hybridization Probes from the Cosmids Containing Multiple Rare Restriction sites

The 49 cosmids containing multiple rare restriction sites were digested with *Eco*RI, *Hin*dIII, or *Pst*I, and the resulting fragments were resolved on 0.8% low-melting-temperature agarose gels. Individual fragments were excised from the gel, radiolabeled, and hybridized to filters to determine whether they were good unique-sequence chromosome 3–specific hybridization probes. Multiple fragments from each cosmid were tested to increase the probability of finding

Table 1

Isolated Cosmids That Contain Multiple Rare Restriction Sites

Cosmid	Chromosome 3 Cluster Cosmids[a]						
	*Bss*HII	*Mlu*I	*Not*I	*Pvu*I	*Sal*I	*Sfi*I	*Sst*II
cA7.....	0	0	0	0	0	1	2
cA8.....	. . .	. . .	0	0	2	1	3
cA13.....	0	0	0	0	0	4	1
cA14.....	1	0	1	0	0	0	2
cA31.....	. . .	. . .	1	0	1	1	3
cA34.....	3	1	2	0	0	2	3
cA41.....	1	1	. . .	. . .	. . .	0	2
cA46.....	2	0	1	0	1	1	2
cA54.....	0	0	1	0	0	2	1
cA70.....	4	2	1	1	0	3	6
cA77.....	2	0	2	0	0	4	3
cA83.....	2	1	1	0	0	3	4
cA84.....	2	0	3	1	0	4	4
cA101	2	1	1	0	1	4	2
cA133	1	0	1	0	0	1	1
cA134	2	0	0	0	2	5	2
cA171	2	2	0	0	2	1	2
cA199	2	0	2	0	0	2	2
cA240	1	1	0	0	0	0	2
cA243	1	1	0	0	0	1	2
cA257	1	1	1	0	0	0	1
cA267	1	1	1	0	0	0	1
cA269	2	0	1	0	1	3	0
cA273	. . .	. . .	0	0	1	1	2
cA278	1	1	1	0	1	1	1
cA298	1	. . .	1	. . .	. . .	0	1
cA322	3	2	1	2	2	3	2
cA327	0	. . .	1	0	1	1	1
cA328	. . .	0	0	0	2	1	2
cA461	0	0	0	0	0	0	2
cA467	1	0	1	1	1	2	1
cA476	1	0	1	1	1	2	5
cA504	1	1	1	0	0	2	1
cA530	. . .	. . .	0	0	0	1	2
cA533	1	0	1	0	0	1	2
cA537	1	0	1	0	0	1	2
cA545	0	0	0	0	1	1	4
cA548	1	. . .	0	0	1	0	2
cA552	1	. . .	1	0	1	0	1
cA556	3	2	2	1	4	1	2
cA562	1	. . .	0	0	2	3	3
cA571	0	1	1	0	0	1	1
cA572	1	0	1	0	0	3	2
cA575	. . .	0	0	0	0	1	3
cA611	1	1	0	1	1	0	2
cA618	. . .	. . .	0	0	0	1	2
cA649	0	0	1	0	1	1	1
cA741	. . .	. . .	0	0	0	0	3
cA764	2	0	2	1	0	0	3

[a] 0 = No discernable change when a single digestion with *Eco*RI is compared with a secondary digestion with one of the rare-cutting restriction endonucleases; . . . = cosmid not tested for sites by using that restriction endonuclease.

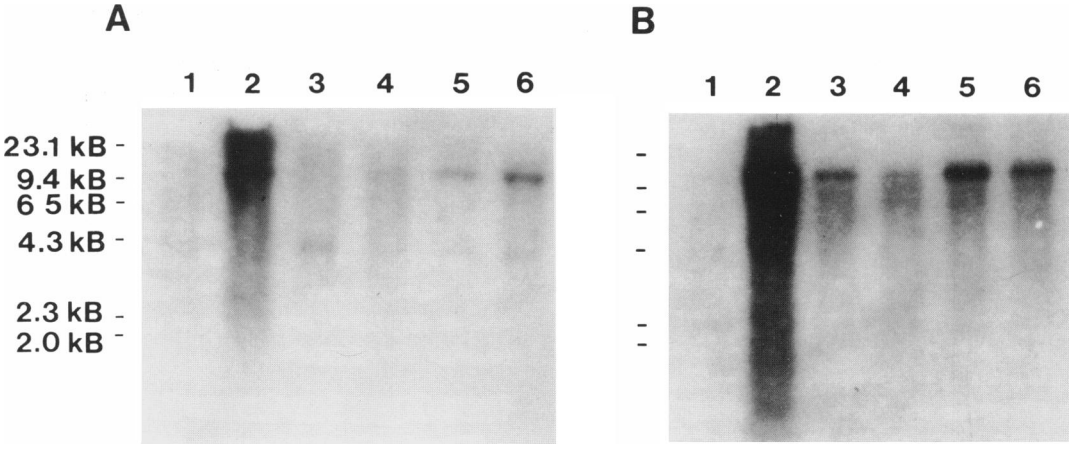

Figure 2 Hybridization of unique-sequence probes from two cluster cosmids to long arm deletion mapping panel. *A*, cA171 *Pst*I + *Sst*II fragment 5 (3q24.2-q26). *B*, cA101 *Hin*dIII fragment 1 (3q26-qter). Lanes 1, CHO. Lanes 2, HeLa. Lanes 3, H3-4. Lanes 4, Hy86D13. Lanes 5, 2A-3 #5. Lanes 6, Hy137RT9. These two probes were apparently evolutionarily conserved, as they hybridize to distinct fragments from hamster and human DNA.

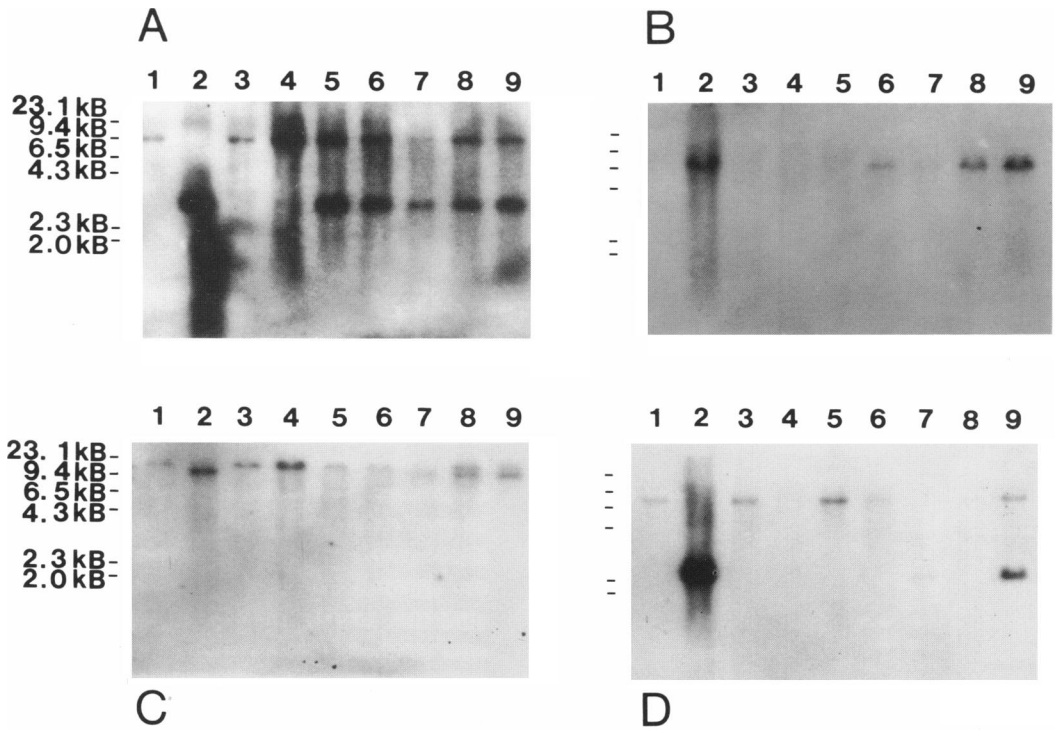

Figure 3 Hybridization of unique-sequence probes from four cluster cosmids to short-arm deletion mapping panel. *A*, cA41 *Eco*RI fragment 3 (3p14.1-14.2). *B*, cA83 *Pst*I fragment 13 (3p14.2-p21.1). *C*, cA13 *Eco*RI fragment 3 (3p21.1-p24.2). *D*, cA8 *Pst*I fragment 14 (3p24.2-pter). Lanes 1, CHO. Lanes 2, HeLa. Lanes 3, UCTP-2A-3 #5. Lanes 4, 182UC/7-3. Lanes 5, TL12-8. Lanes 6, 3;7/UC-1. Lanes 7, R1-1. Lanes 8, 3;21/UC-1. Lanes 9, Hy137RT9. Three of these probes (*A, C,* and *D*) were evolutionarily conserved, as they hybridized to distinct fragments from hamster and human DNA. Other fragments from cA83 were found to be evolutionarily conserved.

several fragments from each cosmid that were good probes. Most of these cosmids were excellent sources of unique-sequence probes, and many of the probes appeared to contain evolutionarily conserved sequences, since they hybridized to both human and CHO sequences.

Subchromosomal Localization of the Cosmids

Several fragments from each cosmid were then hybridized to filters, to more precisely localize each cosmid. We used different sets of filters, depending on whether a fragment was localized to the long or short arm of the chromosome. Figure 2 shows the hybridization of two fragments to the long-arm deletion mapping panel, and figure 3 shows the hybridization of several fragments to the short-arm deletion mapping panel. We tested several fragments from each cosmid to verify the subchromosomal localizations. Figure 4 shows where each of the cosmids was localized on chromosome 3. Seven of the cosmids (cA278, cA322, cA328, cA504, cA572, cA552, and cA556) were not localized with the same degree of resolution as were the others. It was extremely difficult to isolate from these cosmids any fragments which were sufficiently free of repetitive sequences as to be able to precisely localize the cosmids.

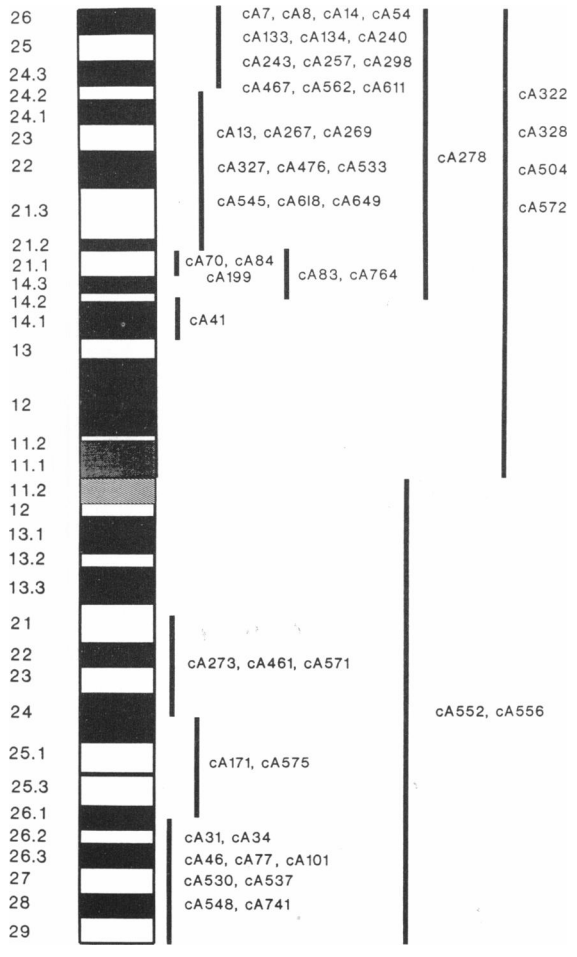

Figure 4 Subchromosomal localization of 49 cluster cosmids. Cosmids cA70, cA84, and cA199 were more precisely localized to 3p21.1 by hybridization to an additional somatic cell hybrid, R10-4 (proximal 3p21.1-qter) (Smith et al. 1989*a*).

Testing Fragments for Sequence Conservation

The initial hybridizations to test fragments derived from these cosmids, as unique-sequence hybridization probes also yielded many fragments that hybridized to both CHO and human DNA. To verify that this hybridization was due to sequence conservation and was not a fragment containing both CHO and human DNA, we tested each fragment that appeared to be conserved on filters with CHO and mouse and UCTP-2A-3 DNAs. Figure 5 shows the results of this type of hybridization with fragments derived from three of the cluster cosmids. This demonstrates that these fragments are evolutionarily conserved. Several of these fragments were hybridized to Poly A^+ and poly A^- mRNA isolated from HeLa cells. Figure 6 shows the result of this hybridization with fragments from four of the cluster cosmids.

Discussion

The goal of the present work was to identify a large number of chromosome 3–specific genes. An analysis of 15% of our currently available resource of chromosome 3–specific cosmids yielded 49 cosmids that contained multiple rare restriction sites. We provide evidence that many of these cosmids may contain gene sequences; such evidence includes both the isolation of evolutionarily conserved fragments from many of these cosmids and the demonstration that several of the cosmids contain transcribed sequences. We are currently testing the remaining chromosome 3–specific cosmids to potentially identify several hundred additional cosmids containing multiple rare restriction sites.

In a previous publication we described the characterization of two chromosome 3–specific cosmids containing multiple rare restriction sites (Smith et al. 1989*a*). Precise restriction maps were generated for

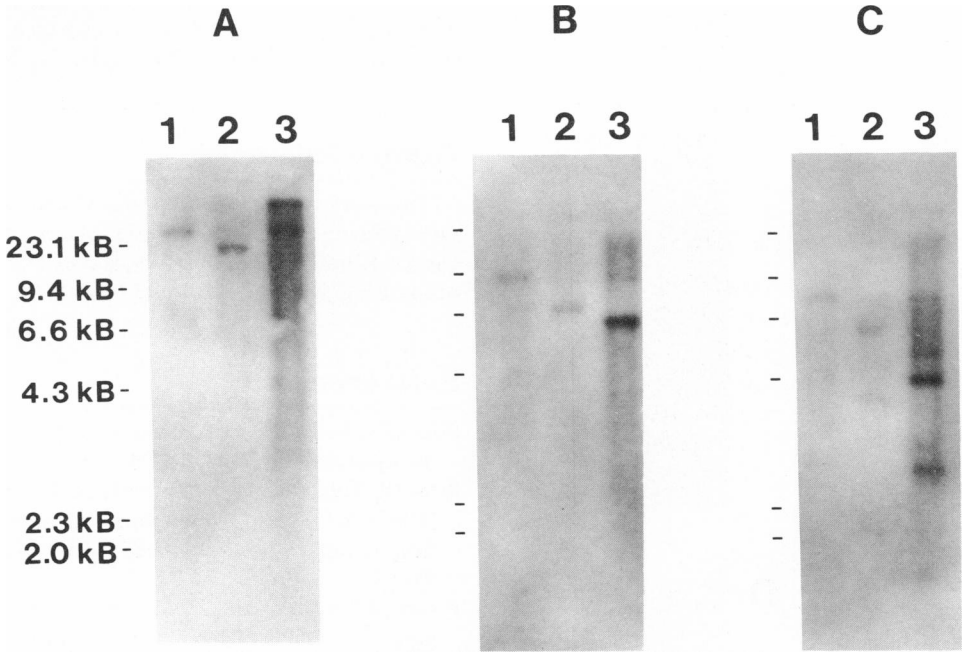

Figure 5 Testing unique-sequence fragments from three cosmids for hybridization to hamster (CHO) (lanes 1), mouse (RAG) (lanes 2), and hybrid (UCTP-2A-3) DNA's (lanes 3). *A*, cA199 *Eco*RI fragment 3. *B*, cA322 *Hin*dIII fragment 7. *C*, cA84 *Not*I + *Sst*II fragment 3.

these cosmids, and we found that the rare restriction sites present within the cosmids were present in one or more tight clusters of rare restriction sites. We have therefore coined the term "cluster cosmids" for cosmids containing multiple clustered rare restriction sites. An important question is whether these cosmids contain classical HTF islands (Bird et al. 1985). To answer this question we have begun to test whether CpG dinucleotides present within tight clusters of rare restriction sites are methylated in a variety of different tissues. Preliminary evidence reveals that the CpGs present within the clusters are indeed unmethylated.

An analysis of the subchromosomal distribution of the cluster cosmids (fig. 4) reveals their nonrandom distribution along the length of the chromosome. The greatest number of cluster cosmids was found in the most telomeric regions (3p24.2-pter and 3q26.2-qter) of the chromosome. Conversely, the centromeric region from 3p14.1-q21 contained no cluster cosmids. Is this a true representation of the subchromosomal distribution of clusters of rare restriction sites on the chromosome? We hope to answer this question after we identify and localize several hundred more cluster cosmids. Are the genes for chromosome 3 similarly clustered in telomeric regions? As of HGM10 there were 63 reported genes on chromosome 3 (Naylor and Bishop 1989). The distribution of the subset of these genes which have been more precisely subchromosomally localized is similar to the distribution of cluster cosmids, but several cloned genes have been localized to more centromeric regions of the chromosome, including GPX1 (McBride et al. 1988) and UMPS (Patterson et al. 1983). It is important to note that the 63 known genes plus the 49 cluster cosmids probably represent only 2% of the total number of chromosome 3 genes.

In the present paper we have demonstrated how easy it is to isolate large numbers of cosmids which are gene candidates. The only prerequisite is to have a large resource of recombinants from one particular chromosome or chromosomal region. As part of the Human Genome Initiative there is a major effort underway to construct cosmid libraries from flow-sorted chromosomes (Deaven and Van Dilla 1990). Once constructed, these libraries (which should be enriched for recombinants from a single human chromosome) can be similarly screened to identify large numbers of cluster cosmids from specific chromosomes.

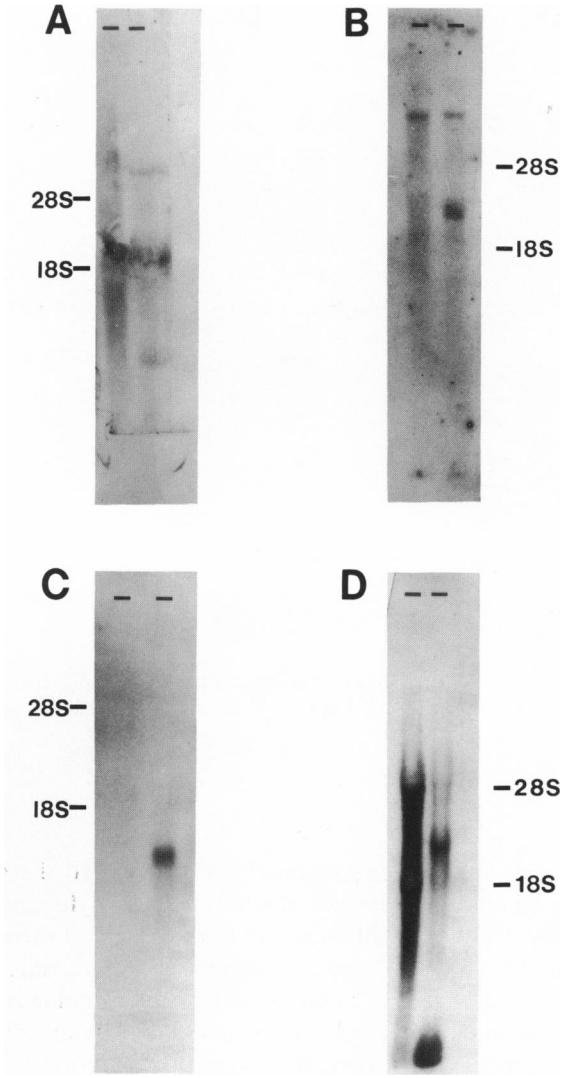

Figure 6 Hybridization of evolutionarily conserved fragments from four cluster cosmids to poly A⁻ and poly A⁺ HeLa mRNA. *A*, cA13 *Eco*RI fragment 3. *B*, cA84 *Bam*HI frgment 6. *C*, cA199 *Hin*dIII fragment 7. *D*, cA54 *Eco*RI fragment 7.

In addition to their usefulness in the identification of genes, the cluster cosmids are also invaluable mapping tools. Since the cluster cosmids have multiple rare restriction sites, and since these sites are generally clustered and unmethylated, they can function as large linking clones. Two of the first cluster cosmids identified by us—cA84 and cA199 (Smith et al. 1989*a*)—have facilitated the construction of a precise macrorestriction map spanning 2.5 mB within 3p21.1 (Gemmill et al., in press). We are currently attempting to expand the macrorestriction map by analyzing cluster cosmids localized within 3p14.2-p24.2.

Acknowledgments

This work was supported by the Molecular Biology Center at Wayne State University, by March of Dimes—Birth Defects Foundation grant 5-645, and by National Institutes of Health grants CA48045 and HG00245.

References

Bird AP (1986) CpG-rich islands and the function of DNA methylation. Nature 321:209–213

Bird AP, Taggart M, Frommer M, Miller OJ, MacLeod D (1985) A fraction of the mouse genome that is derived from islands of nonmethylated CpG-rich DNA. Cell 40: 91–99

Brown WRA, Bird AP (1986) Long-range restriction mapping of mammalian genomic DNA Nature 322:477–481

Carlock LC, Smith DI, Wasmuth JJ (1986) Genetic counterselective procedure to isolate interspecific hybrids containing single human chromosomes: construction of cell hybrids and recombinant DNA libraries specific for human chromosomes 3 and 4. Somatic Cell Mol Genet 12: 163–174

Deaven LL, Van Dilla MA (1990) National Laboratory Gene Library Project: construction of DNA libraries from flow-sorted chromosomes. Hum Genome News 2:12–13

Drabkin HA, LeBeau MM, Rowley JD, Glode LM, Carino J, Diaz M (1987) Isolation and analysis of the acute myelogenous leukemia (3;3) translocation. Proc Am Assoc Cancer Res 28:A202

Drabkin H, Jonsen M, Wright M, Varkony T, Jones C, Sage M, Gold S, et al (1990) Development of a somatic cell hybrid mapping panel and molecular probes for human chromosome 3. Genomics 8:435–446

Favaloro J, Treisman R, Kamen R (1980) Transcription maps of polyoma virus–specific RNA: analysis by two dimensional nuclease S1 gel mapping. Methods Enzymol 65:718

Feinberg AP, Vogelstein B (1983) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal Biochem 132:6–13

Gemmill RM, Garcia M, Smith DI, Erickson P, Miller Y, Coyle-Morris J, Tommerup N, et al. A 2.5 Mb physical map within 3p21.1 detects the breakpoint associated with Greig cephalopolysyndactyly syndrome. Genomics (in press)

Lindsay S, Bird AP (1987) Use of restriction enzymes to detect potential gene sequences in mammalian DNA. Nature 327:336–338

McBride OW, Lee BJ, Hatfield DL, Mullenbach G (1988)

Glutathione peroxidase gene maps on human chromosomes 3, 21 and X. FASEB J 2:A1001

McMaster GK, Carmichael GG (1977) Analysis of single and double stranded nucleic acids on polyacrylamide and agarose gels by using glyoxal and acridine orange. Proc Natl Acad Sci USA 74:4835–4839

Mears JG, Ramirez F, Leibowitz D, Nakamura F, Bloom A, Kanotey-Ahulu F, Bank A (1978) Changes in restricted human cellular DNA fragments containing globin gene sequences in thalassemias and related disorders. Proc Natl Acad Sci USA 75:1222–1226

Naylor SL, Bishop DT (1989) Report of the Committee on the Genetic Constitution of Chromosome 3 (HGM10). Cytogenet Cell Genet 51:106–120

Patterson D, Jones C, Morse H, Rumsby P, Miller Y, Davis R (1983) Structural gene coding for multifunctional protein carrying orotate phosphoribosyltransferase and OMP decarboxylase activity is located on long arm of human chromosome 3. Somatic Cell Genet 9:359–374

Smith DI, Golembieski W, Drabkin H, Kiousis S (1989*a*) Identification of two cosmids derived from within chromosomal band 3p21.1 that contain clusters of rare restriction sites and evolutionarily conserved sequences. Am J Hum Genet 45:443–447

Smith DI, Golembieski W, Gilbert JD, Kizyma L, Miller OJ (1987) Overabundance of rare-cutting restriction endonuclease sites in the human genome. Nucleic Acids Res 15:1173–1184

Smith DI, Mangrulker R, Geist R, Gilbert J, Kinsman K, Drabkin H, Golembieski W (1989*b*) Saturation of human chromosome 3 with unique sequence hybridization probes. Genomics 4:453–459