Cloning of 559 Potential Exons of Genes of Human Chromosome 21 by Exon Trapping

Haiming Chen,[1] Roman Chrast,[1] Colette Rossier,[2] Michael A. Morris,[2]
Maria D. Lalioti,[1] and Stylianos E. Antonarakis[1–3]

[1]Laboratory of Human Molecular Genetics, Department of Genetics and Microbiology, Geneva University Medical School, and [2] Division of Medical Genetics, Cantonal Hospital of Geneva, Switzerland

Chromosome 21 represents ~1% of the human genome, and its long arm has been estimated to contain 600–1000 genes. A dense linkage map and almost complete physical maps based on yeast artificial chromosomes (YACs) and cosmids have been developed. We have used exon trapping to identify portions of genes from randomly picked chromosome 21-specific cosmids, to contribute to the creation of the transcription (genic) map of this chromosome and the cloning of its genes. A total of 559 different sequences were identified after elemination of false-positive clones and repetitive elements. Among these, exons for 13 of the 30 known chromosome 21 genes have been "trapped." In addition, a considerable number of trapped sequences showed homologies to genes from other species and to human expressed sequence tags (ESTs). One hundred thirty-three trapped sequences were mapped, and every one mapped back to chromosome 21. We estimate that we have identified portions of up to ~40% of all genes on chromosome 21. The genic map of chromosome 21 provides a valuable tool for the elucidation of function of the genes and will enhance our understanding of the pathophysiology of Down syndrome and other disorders of chromosome 21 genes.

The cloning of human genes and elucidation of the function of their protein products is of fundamental importance for the understanding of the etiology and pathophysiology of human hereditary disorders. One of the goals of the international effort known as the Human Genome Project is to identify, map, and determine the nucleotide sequences of all human genes (Collins and Galas 1993). The priority of determining the nucleotide composition of protein-coding portions of the genome is justified by the medical relevance of this information with regard to both monogenic disorders and the common disease phenotypes including neoplasias.

Chromosome 21 is the smallest human chromosome, the long arm of which has been estimated to comprise ~1% of the human genome (Antonarakis 1993). The total number of genes on this chromosome is predicted to be ~600–1000 (for the estimation of the total number of human genes, see Fields et al. 1994). The linkage map of chromosome 21 is one of the most dense of all human chromosomes, with more than 120 highly polymorphic short sequence repeats mapped in a total sex-averaged length of 67 M (McInnis et al. 1993; Antonarakis et al. 1995; A. Chakravarti and S.E. Antonarakis, in prep.). The physical contig of the 38 Mb of the long arm of chromosome 21, using yeast artificial chromosomes (YACs) and other cloning systems [cosmids, P1s, P1 artificial chromosomes (PACs), bacterial artificial chromosomes (BACs)], is almost complete (Chumakov et al. 1992; Nizetic et al. 1994). Only 30 known chromosome 21 genes have been cloned and sequenced to date (Genome DataBase search on June 28, 1995). We have used exon trapping (Buckler et al. 1991; Church et al. 1994) to identify portions of genes and to contribute to the development of the complete transcription (genic) map of this chromosome and thereby to the understanding of the etiology of the phenotypes of Down syndrome and other disorders involving chromosome 21 genes. We report here the cloning, sequencing, and partial characterization of DNA sequences that represent portions of up to ~40% of the predicted number of genes on human chromosome 21. Further study of the cDNAs corresponding to the trapped exons will enhance our understanding of chromosome 21-related disorders and the

[3]Corresponding author.
E-MAIL sea@medsun.unige.ch; FAX 41-22-702-5706.

6:747–760 ©1996 by Cold Spring Harbor Laboratory Press ISSN 1054-9803/96 $5.00

role of this chromosome in normal human development and physiology.

RESULTS

The method of exon trapping (Buckler et al. 1991; Church et al. 1994) was used to identify portions of human genes that map on chromosome 21. Cosmids taken at random from the chromosome 21-specific library LL21NCO2-Q (Soeda et al. 1995) were used for identification of genes throughout the entire chromosome. A total of 1194 cosmids were used. Pools of 10 cosmids were used for each trapping experiment. In only a few experiments, all cosmids from a 96-microliter well plate were used. Clones that contained human ribosomal RNA (RNR) sequences and mouse genomic sequences were eliminated and not used for trapping (see Methods). Because the average length of the inserts was ~40kb, the 1194 cosmids represent ~48 Mb of chromosome 21 DNA, which is similar to the estimated size of both arms of this chromosome (Ichikawa et al. 1993). After elimination of false-positive exons caused by vector self-splicing events (see Methods), a total of 1030 potential exons were trapped and sequenced (Table 1). Of these, 619 were unique sequences, whereas the remaining 411 were redundant. Fifty-five different trapped sequences showed homology to highly or moderately repeated humam genomic elements (including Alu, LINE, MER1 repeats, and RNR genes; see Table 1); five clones contained contaminant *Escherichia coli* sequences. After elimination of these 60 sequences, a total of 559 different trapped sequences were identified. The complete nucleotide sequences of all of these potential exons have been deposited in EMBL/GenBank (accession nos. X88001–X88560, X86349–X86351, X83219, X84366, and X83513–X86516).

The size distribution of the different trapped sequences is shown in Figure 1. The mean size of the trapped exons was 125 nucleotides with a standard deviation of 60 nucleotides; the median size was 115 nucleotides. The GC content of the trapped sequences is similar to that of cDNAs and distinctly higher than that of genomic sequences. The GC content of 1114 kb of genomic sequences was 42.5%, whereas that of 259 kb of cDNAs was 49.6% (sequences were selected randomly from the GenBank/EMBL data bases). The GC content of the trapped sequences reported here (total sequence length of 65.4 kb) was 51.4%.

A total of 30 exons (5.4% of the 559 different sequences) were identical to exons of 13 genes identified previously, known to map on human chromosome 21. These homologies are shown in Table 2.

Table 3 shows the findings of the remaining homology searches. We used the probability of 10^{-4} as a cutoff point for significance in the homology searches. Using this criterion, a total of 378 (67.6% of 559) of potential exons did not show significant homologies to existing entries in the nucleotide and protein data bases. Fifty-three sequences (9.5% of 559) showed identity or strong homology to human expressed sequence tags (ESTs) (Table 3F). The predicted translation products of 83 trapped sequences (14.8% of 559) showed a considerable degree of homology to proteins from the data bases. The homology was convincing in 49 (Table 3A), but weaker in 21, of those exons (Table 3D); 9 had homologies to the collagen gene families (Table 3B), and 7 to Pro- or Cys-rich proteins (Table 3C). Some of the outstanding homologies include those to predicted polypeptides of genes for *Drosophila single-minded, white,* and *enhancer of zeste,* rat lanosterol synthase, and megalin, bovine ATP synthase OSCP subunit, yeast PWP2 and one protein kinase, *Xenopus* neural cell-adhesion molecule, mouse pericintrim, T-cell invasion and metastasis protein, requiem, human coagulation factor 11, and elastase 2b (see corresponding GenBank accession nos. for their references). Three further sequences were identical to cloned but unmapped human genes such as the *GABPA* transcription factor (Watanabe et al. 1993) and members of the β*2-chimerin* gene family (Leung et al. 1994) (Table 3A, sequences 1–3). Some clones (Table 3F, e.g., sequences 133–139) showed identity to areas of chromosome 21 that have been sequenced as part of the cosmid sequencing project at the Lawrence Berkeley laboratory (C.H. Martin, M.M. Bondoc, A. Chiang, T. Cloutier, C.A. Davis, C.L. Ericsson, M.A. Jaklevic, R.J. Kim, M.T. Lee, M. Li, C.A. Mayeda, A. Steiert-El Kheir, and M.J. Palazzolo, unpubl.; GenBank accession no. L35676).

A subset of the trapped exons have been mapped back to chromosome 21 using different methods. To date, not a single trapped exon tested has been mapped in a genomic region outside of human chromosome 21. A total of 133 exons have been mapped to chromosome 21 by (1) hybridization or PCR amplification using chromosome 21 cosmids (67 exons), YACs (35

Table 1. Results of the Exon Trapping Experiment Using Randomly Picked Chromosome 21-specific Cosmids.

	Number (%)	Duplicates
Total clones sequenced[a]	1030	
Different trapped sequences[a]	619	(+411)
Different trapped sequences[b]	559	(+391)
Known genes on HC21(Table 2)		
known HC21 genes hit	13	
different trapped sequences	30 (5.4% of 559)	(+43)
Excellent homologies (Table 3) up to $P < 10^{-4}$	83 (14.8% of 559)	(+71)
convincing homologies to genes	49 (8.7% of 559	(+43)
weaker homologies to genes	21 (3.7% of 559)	(+21)
collagen gene family homologies[c]	9 (1.6% of 559)	(+1)
pro- or cys-rich protein homologies	7 (1.2% of 559)	(+6)
Known but unmapped human genes	3 (0.5% of 559)	(+3)
ESTs	53 (9.5% of 559)	(+45)
"Minus" strand homologies	5 (0.9% of 559)	(+1)
New trapped sequences[d]	526 (94.1% of 559)	(+46)
Repetitive elements	55 (8.9% of 619)	(+20)
RNR genes	4	
Alu element	33	
LINE	3	
MER repeat	4	
*Mst*II repeat	2	
pericentric 48 bp repeat	1	
THE-1 element	1	
O family repeat	1	
α satellite	1	
TR7 repeat	2	
SST repeat	1	
THR repeat	1	
Contaminants (*E. coli* sequences)	5	
Total trapped sequences mapped to HC21	133 of 133 (23.8% of 559)	

[a] Excluding the false positive (vector self-splicing), no insert clones or poor quality of sequence.

[b] Excluding repetitive elements and *E. coli* contaminants.

[c] These include homologies with collagen motifs (G*xx*G*xx*) and glycine-rich sequences.

[d] Includes all different trapped sequences (b) except those identical to the known chromosome 21 genes and the known but unmapped genes.

exons), and rodent–human somatic cell hybrids (40 exons); (2) because they were identical to chromosome 21 genes identified previously (30 exons); (3) because they were identical to sequences from chromosome 21 (39 exons). A few exons were additionally mapped by fluorescence in situ hybridization (FISH) analysis using their corresponding cosmids as probes. Some exons were mapped to chromosome 21 by more than one means (Table 3).

To verify whether the trapped sequences contain parts of genes, we performed cDNA library screening using pools of these sequences as probes. Two pools of 50 trapped sequences were hybridized to 300,000 plaques of two cDNA libraries: Pool 1 was hybridized to an amplified retina cDNA library (Nathans et al. 1986), and pool 2 to a commercially available fetal brain cDNA library (Clontech). A total of 126 and 29 strongly positive plaques were identified, respectively, after overnight autoradiography. Two complex probes from 20 and 23 plaque-purified clones were made by PCR amplification from pooled DNA from these clones, and hybridized against the original 100 trapped sequences. A total of 5 and 12 trapped sequences were positive

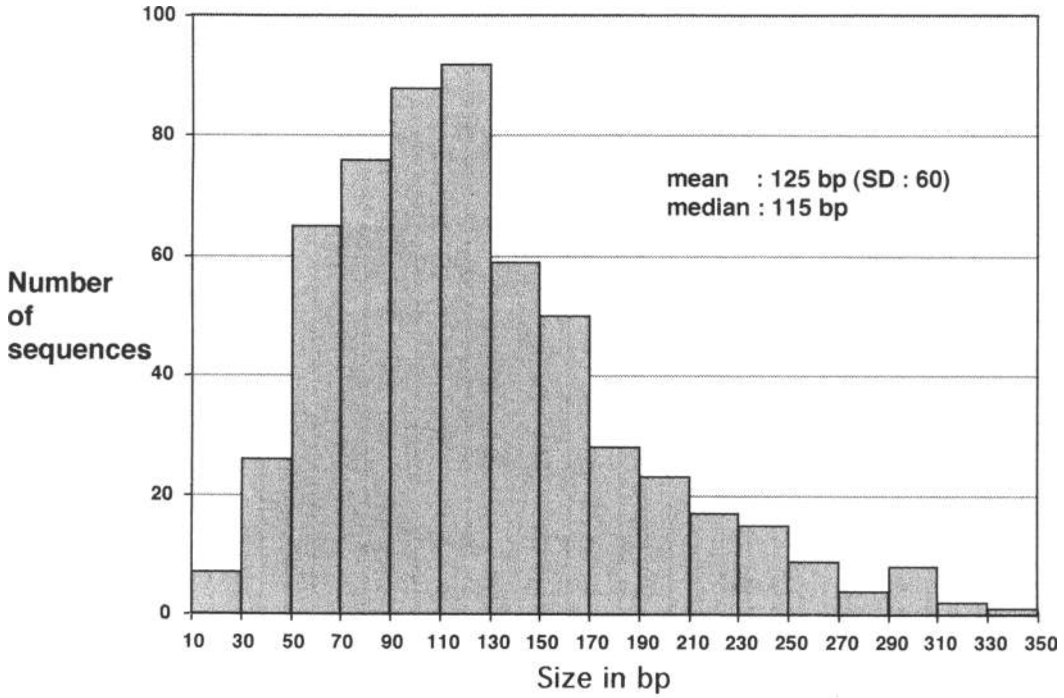

Figure 1 Histogram of sizes (in nucleotides) of the 559 different exon-trapped sequences.

from pools 1 and 2, respectively. As 20 of 126 and 23 of 29 strongly positive plaques were used to generate the probes for back-hybridization, these data imply that at least 31 and 15 trapped sequences from pools 1 and 2, respectively, are present in these cDNA libraries (62% and 30% of trapped clones). It should be noted that these are underestimates because (1) pooled probes were used (selecting against small or poorly labeled exons); (2) only strongly positive cDNA clones were picked; (3) the back-hybridization probes were generated by whole-insert PCR (selecting against large inserts); and (4) cDNA libraries from only two tissues were tested.

DISCUSSION

We have used exon trapping to clone gene fragments and contribute to the transcription (genic) map of chromosome 21. This method was chosen because it is independent of the tissue-specific, temporal, and spatial expression of genes and does not rely on the abundance of a particular clone in a given cDNA library (Buckler et al. 1991; Brennan and Hochgeschwander 1995). Targeted experiments with a cosmid containing the entire *PFKL* gene (the sequence of its exons and splice juctions were known) (Elson et al. 1990) showed that the method captures genomic portions demarkated by donor and acceptor splice sites; although cryptic splice sites can be used in the experimental strategy, the majority of the trapped sequences contained at least one authentic splice site of exons (data not shown).

Pools of randomly picked chromosome 21 cosmids were used to trap the exons. We estimate that we have identified portions of ≤40% of chromosome 21 genes for two reasons: (1) As expected, a number of exons from the chromosome 21 genes cloned previously were also identified by our whole-chromosome approach. Exons from a total of 13 genes identified previously have been sequenced (Table 2). Because nucleotide sequences of 30 chromosome 21 genes with more than two exons are included in the sequence data bases, we conclude that we have probably captured ~40% of the genes of this chromosome. (2) For these 13 known chromosome 21 genes, a total of 30 different exons have been captured (one known exon for every 19 trapped different sequences) of 2.3 exons per known gene. If the same numbers apply to the remainder of the trapped exons, the 526 additional different captured sequences correspond to ~230 genes. If the total number of genes on the long arm of human chromosome 21 (which is

Table 2. Trapped Exons Identical to Portions of Chromosome 21 Genes

	Clone ID	GenBank	Size	BLASTN	BLASTX
1	hmc20f08	X88307	226	APP X06989 nt 1755-1976 4.4e-83	APP P05067 aa 564-636 1.0e-40
2	hmc20f07	X88308	58	APP X06989 nt 1977-2030 2.6e-14	APP P05067 aa 638-654 1.6e-4
3	rmc07f11	X88557	151	APP X06989 nt 2132-2278 1.4e-53	APP P05067 aa 689-737 3.9e-26
4	hmc27g05	X88209	210	IFNRAF-1 U05877 nt 320-525 1.9e-76	IFNRAF-1 P38484 aa 70-137 9.2e-44
5	hmc16f08	X88361	44	AML1 U19601 nt 70-98 6.5e-4	AML1 U19601 aa 24-33 9.5e-1
6	hmc16b10	X88368	161	AML1 U19601 nt 352-508 2.3e-55	AML1 U19601 aa 118-169 9.3e-27
7	hmc09d06	X88431	165	AML1 U19601 nt 806-926 1.3e-55	AML1 U19601 aa 243-282 3.0e-28
8	hmc24c11	X88256	90	CBR M62420 nt 483-567 1.4e-26	CBR P16152 aa 69-95 7.4e-11
9	hmc48a01	X88010	112	CBR M62420 nt 1112-1219 3.3e-36	CBR P16152 aa 98-131 9.7e-18
10	hmc21c04	X88297	156	ERG M17254 nt 514-668 1.7e-53	ERG P11308 aa 87-137 4.1e-30
11	hmc48g10	X88004	73	ERG M17254 nt 951-1019 1.5e-20	ERG P11308 aa 233-254 5.0e-10
12	hmc28f06	X88194	240	ERG M17254 nt 1020-1124 6.7e-35	ERG P11308 aa 257-289 2.2e-16
13	hmc14g09	X88379	116	ETS2 J04102 nt 364-477 8.3e-37	ETS2 P15036 aa 25-62 2.9e-16
14	hmc42f12	X88063	123	ETS2 J04102 nt 1367-1485 7.8e-41	ETS2 P15036 aa 360-398 1.4e-21
15	hmc04e09	X88484	111	Enterokinase U09860 nt 814-920 6.3e-37	Enterokinase U09860 aa 259-293 1.9e-17
16	hmc06h02	X88464	197	MX1 M30817 nt 316-508 3.5e-69	MX1 P20591 aa 35-98 4.0e-32
17	hmc30d07	X88179	142	MX1 M30817 nt 509-649 1.1e-48	MX1 P20591 aa 100-144 1.8e-22
18	hmc43b05	X88062	139	MX2 M33883 nt 360-497 3.9e-40	MX2 P20592 aa 149-192 1.4e-25
19	hmc27f09	X88213	206	MX2 M30818 nt 1175-1377 8.6e-74	MX2 P20592 aa 358-424 1.9e-35
20	hmc12d01	X88411	141	EHOC-1 U19252 nt 287-422 7.5e-44	EHOC-1 U19252 aa 51-95 1.5e-22
21	hmc41e10	X88070	200	EHOC-1 U19252 nt 620-815 1.9e-71	EHOC-1 U19252 aa 162-226 2.2e-37
22	hmc02a10	X88510	120	EHOC-1 U19252 nt 1323-1439 2.6e-31	EHOC-1 U19252 aa 396-433 2.5e-12
23	hmc26a11	X88225	161	CD18 M15395 nt 70-177 1.7e-32	CD18 P05107 aa 1-34 2.9e-12
24	hmc25e08	X88236	131	CD18 M15395 nt 687-813 1.2e-40	CD18 P05107 aa 207-247 1.5e-18
25	hmc19a05	X88322	67	COL18A1 L22548 nt 662-724 7.4e-18	COL18A1 P39060 aa 222-241 1.7e-8
26	hmc21b09	X88298	134	COL6A1 X15879 nt 146-275 4.1e-46	COL6A1 S05337 aa 34-76 31.2e-22
27	hmc21a12	X88300	205	COL6A1 X15879 nt 276-476 1.9e-76	COL6A1 S05337 aa 77-143 1.7e-42
28	hmc21e04	X88291	131	COL6A1 M20776 nt 37-147 6.2e-31	COL6A1 S05337 aa 269-304 1.1e-17
29	hmc04c10	X88488	181	COL6A1 M20776 nt 631-807 1.1e-62	COL6A1 S05377 aa 467-525 1.1e-39
30	hmc21g06	X88286	24	COL6A1 X15879 nt 276-299 7.8e-2	short seq to show homology

In the BLAST homology columns the following information is included: gene name, GenBank sequence accession no., region of nucleotide (nt) or amino acid (aa) homology, and *P* value.

~1% of the human genome) is between 600–1000, then we estimate that the unknown exons described here represent portions of 23%–38% of the genes on chromosome 21.

A total of 1194 cosmids were used in the experiments described here and 559 different exons have been identified, representing an average of 0.47 exons per cosmid used, or one exon per 85 kb of DNA. These numbers are not different from the experience of other investigators who have used exon trapping with similar numbers of cosmids for each pool (Buckler et al. 1991; Church et al. 1993). The continuation of the exon trapping experiment with ~1500 more cosmids will result in the cloning of exons from perhaps as much as 80%–90% of the genes on chromosome 21. It is difficult, however, to achieve a more complete coverage of the transcription (genic) map before a cosmid contig has been determined, and all the captured exons have been mapped back to their corresponding cosmid clones. Cosmids from this contig with no corresponding exons, or cosmids not used previously for exon trapping could then be used for directed completion of the transcription map. Futhermore, because exon trapping does not identify gene portions from genes with less than three exons (exons without both donor and acceptor splice sites), other methods should be used for the identification of such genes. For example, the genes for Na$^+$/myo-inositol cotransporter (Berry et al. 1995) and Isk potassium channel (Murai et al. 1989) are intronless, and therefore they could not have been identified by the strategy described here. Other methods, such as 3′ exon trapping (Krizman and Berget 1993), or cDNA selection (Parimoo et al. 1991; Lovett et al. 1991), among others will therefore be required for the completion of the transcription map. Three studies using cDNA selection that identified a number of cDNAs from certain regions of

Table 3. Trapped Exons from HC21-specific Cosmids; Homologies to Sequences in the Data Bases

		Other Homologies						Mapping
		Clone ID	GenBank	Size	BLASTN	BLASTX	HC21[a]	mode[b]
A	1	rch04a07	X84366	107	human E4TF1-60 D13318 nt 167-269 2.4e-35	human E4TF1-60 A48146 aa 1-25 2.3e-10	YES	c,y,h,f
	2	hmc02b06	X88506	61	human beta2-chimaerin U07223 nt 527-582 5.7e-12	human beta2-chimaerin A53764 aa 29-46 3.9e-5	YES	c,y,f
	3	hmc02f04	X88502	36	human beta2-chimaerin U07223 nt 583-614 1.0e-05	human beta2-chimaerin A53764 aa 47-57 1.8e-1	nd	
	4	hmc05d12	X88471	142	no homology	xenopus NCAM1 P16170 aa 162-206 5.7e-15	YES	c,y,h,f
	5	hmc16g08	X88357	79	chicken NCAM2 M15922 nt 17-87 1.7e-6	mouse NCAM3 P13594 aa 20-41 8.2e-2	YES	c,y,h,f
	6	hmc13f06	X83514	93	Drosophila SIM M19020 nt 202-284 2.3e-11	drosophila SIM P05709 aa 86-115 2.7e-13	YES	c,y,h,f
	7	hmc29c01	X83515	113	no homology	drosophila SIM P05709 aa 116-150 4.7e-9	YES	c,y,h,f
	8	hmc05f04	X83513	90	Drosophila SIM M19020 nt 481-561 1.3e-12	drosophila SIM P05709 aa 179-205 8.8e-9	YES	c,y,h,f
	9	rmc03d09	X83516	111	Drosophila SIM M19020 nt 770-872 2.0e-8	drosophila SIM P05709 aa 273-307 2.2e-7	YES	c,y,h,f
	10	hmc21g10	X88284	243	new	drosophila WHITE P10090 aa 270-347 1.1e-17	YES	c,h
	11	hmc21a05	X88301	160	EST F06912 1.7e-42	drosophila WHITE P10090 aa 572-611 1.3e-8	YES	c,h
	12	hmc21c05	X88296	106	mouse ABC8 (white) Z48745 nt 1231-1332 6.9e-28	mouse ABC8 (white) Z48754 aa375-408 1.9e-15	YES	c,h
	13	hmc23b04	X88270	118	EST H75287 yu58h12.r1 1.6e-20	drosophila E(z) U00180 aa 665-694 7.6e-11	YES	c,y,h
	14	hmc17c03	X88345	225	new	drosophila sperm protein Q01643 aa 26-52 3.6e-11	nd	
	15	hmc17b09	X88347	229	EST R00718 ye74d12.r1 1.0e-44	drosophila cAMP dep Pdiesterase P12252 aa 210-274 4.4e-7	nd	
	16	hmc08d05	X83219	247	bovine ATPO M18753 nt 268-514 3.0-e67	bovine ATPO P13621 aa 67-147 1.7e-41	YES	c,y,h,f
	17	hmc47c10	X88012	117	bovine ATPO M18753 nt 398-514 1.6e-25	bovine ATPO P13621 aa 111-147 2.0e-15	YES	c,y,h,f
	18	hmc14f01	X88394	280	rat lanosterol synthase D45252 nt 497-717 3.5e-66	rat lanosterol synthase D45252 aa 146-216 4.0e-42	YES	c,y,h
	19	hmc13h11	X88398	101	rat lanosterol synthase D45252 nt 619-717 5.8e-25	rat lanosterol synthase D45252 aa 186-216 1.3e-13	YES	c,y,h
	20	hmc14b03	X88389	113	rat lanosterol synthase D45252 nt 852-964 5.8e-23	rat lanosterol synthase D45252 aa 263-298 2.6e-16	YES	c,y,h
	21	hmc14b02	X88386	162	rat lanosterol synthase D45252 nt 1189-1334 1.0e-27	rat lanosterol synthase D45252 aa 374-423 1.3e-18	YES	c,y,h
	22	hmc42b04	X88069	59	no homology	rat lanosterol synthase D45252 aa 673-690 2.6e-4	YES	c,y,h
	23	hmc42b10	X88067	83	rat lanosterol synthase D45252 nt 2057-2135 2.4e-13	rat lanosterol synthase D45252 aa 665-690 1.5e-10	YES	c,y,h
	24	hmc28g07	X88191	146	rat lanosterol synthase D45252 nt 2136-2274 3.7e-29	rat lanosterol synthase D45252 aa 691-732 1.6e-18	YES	c,y,h
	25	hmc44e11	X88043	124	human L35682 nt 1095-1214 (-) 4.0e-41	rat megalin L34049 aa 1313-1346 1.2e-5	YES	s,h
	26	hmc17c09	X88343	344	EST R20872 yg05h01.r1 1.3e-63	yeast PWP2 P25635 aa 580-652 6.1e-15	YES	c,y,h
	27	hmc18f10	X88330	291	human S/T protein kinase Z25423 nt 1-100 6.7e-32	yeast protein kinase P14680 aa 439-522 2.4e-24	YES	c,y,h
	28	hmc18a08	X88327	97	rat Yak1 kinase X79769 nt 364-456 4.6e-25	rat Yak1 kinase X79769 aa 79-109 8.2e-12	YES	c,y,h
	29	hmc27g09	X88208	130	rat Yak1 kinase X79769 nt 1676-1777 4.6e-26	rat Yak1 kinase X79769 aa 517-557 5.2e-17	YES	c,y,h
	30	hmc02a08	X88511	233	yeast Z28201 nt 625-740 9.0e-6	yeast protein 64 kD P36043 aa 399-459 6.0e-7	nd	
	31	hmc06a02	X88468	137	new	yeast ATP-dep permease P25371 aa 445-485 2.2e-5	nd	
	32	hmc19a07	X88321	290	human mx1 region L35676 nt 1823-1956 & 2291-2445	6 human coag F11 P03951 aa 572-618 1.0e-16	YES	s
	33	hmc26a01	X88229	220	dog protease M24665 191-361 2.3e-11	human elastase 2b P08218 aa 16-75 1.9e-15	nd	
	34	hmc20c05	X86351	177	mouse TIAM-1 U05245 nt 1918-2093 3.8e-51	mouse TIAM-1 A54146 aa 472-529 5.9e-35	YES	c,y,h
	35	hmc20a04	X86350	151	mouse TIAM-1 U05245 nt 2502-2648 2.1e-40	mouse TIAM-1 A54146 aa 666-714 1.0-23	YES	c,y,h
	36	hmc17f08	X86349	178	mouse TIAM-1 U05245 nt 4641-4812 6.8e-46	mouse TIAM-1 A54146 aa 1379-1435 5.8e-30	YES	c,y,h

Table 3. (*Continued*)

		Other Homologies						Mapping
		Clone ID	GenBank	Size	BLASTN	BLASTX	HC21[a]	mode[b]
	37	hmc46b12	X88027	173	mouse Pericentrin U05823 nt 342-481 5.8e-35	mouse Pericentrin A53188 aa 17-61 3.8e-13	YES	c,h,f
	38	hmc22g01	X88276	117	mouse Pericentrin U05823 nt 1257-1368 1.4e-22	mouse Pericentrin A53188 aa 322-358 2.5e-12	YES	c,h,f
	39	hmc25g07	X88231	159	mouse Pericentrin U05823 nt 6282-6378 7.9e-6 3UTR	no homology	YES	c,h,f
	40	hmc18h10	X88329	201	mouse PEP-19 S65225 nt 1171-1323 1.8e-35	no homology	YES	c,y,h
	41	hmc06b05	X88467	162	new	mouse Ca ph/ase CNCM Q01065 aa 52-103 5.7e-8	nd	
	42	hmc46d05	X88024	119	EST T64415 yc48e09.s1 6.4e-4	mouse requiem U10435 aa 316-351 7.1e-6	nd	
	43	hmc24a02	X88259	172	new	drosophila ovarian tumor locus P10383 aa 520-541 7.4e-5	nd	
	44	hmc37a09	X88106	81	EST D31072 fetal lung 1.2e-23	rape homeodomain protein S41980 aa 245-268 7.6e-3	YES	c,y,h
	45	hmc24b08	X88258	181	EST N46140 yy37d03.r1 3.0e-58	c. elegans F54E7.7 gene U00067 aa 60-106 5.8e-6	nd	
	46	hmc26e04	X88219	260	new	human nitric oxide synthase aa 1227-1252 4.6e-5	nd	
	47	hmc26a05	X88228	165	human mx1 region L35680 nt 747-871 3.8e-39	sea anemone GLHR P35409 aa 148-167 5.8e-4	YES	s
	48	hmc23h02	X88260	59	EST R91742 yp98g05.r1 6.4e-15	zebrafish es1 protein U10403 aa 155-171 1.2e-2	nd	
	49	hmc28b06	X88203	59	EST Z47290 21fi8l23 4.6e-13	yeast Yel029p U18530 aa 29-44 5.4e-2	YES	c,y
B	50	hmc04a06	X88492	91	new	mouse col18a1 L16898 aa 210-237 6.8e-8	nd	
	51	hmc16c07	X88366	277	new	sponge colf1 S31521 1.5e-6	nd	
	52	hmc17h04	X88336	303	new	rat col1a1 P02454 4.9e-7	nd	
	53	hmc31c11	X88175	274	new	chicken col12a1 5.9e-5	nd	
	54	hmc25e03	X88238	216	new	mouse col3a1 X52046 9.7e-5	nd	
	55	hmc28f10	X88193	166	new	c.elegans coldpy13 P17657 3.2e-6	nd	
	56	hmc33c05	X88149	165	new	ascaris cola4 B44982 3.3e-5	nd	
	57	hmc35g12	X88118	313	new	mouse col12a1 C44479 1.1e-7	nd	
	58	hmc36a09	X88115	171	new	c.elegans col8 P18833 1.9e-5	nd	
C	59	hmc10a08	X88427	>300	new	wheat gliadin M11336 4.6e-8 Pro-rich	nd	
	60	hmc12b05	X88414	218	new	arabidopsis ATAPG-1 S21961 8.8e-8 Pro-rich	nd	
	61	hmc32c02	X88165	234	lymnea conopressin M86610 nt 1533-1647 1.4e-5	human U1snRNP M18465 6.0e-6 Pro-rich	nd	
	62	hmc36d07	X88112	197	new	human phosphoprotein B27307 4.4e-5 Pro-rich	nd	
	63	hmc40f06	X88079	221	new	human prpL2 X86019 1.5e-5 Pro-rich	nd	
	64	hmc45e10	X88036	169	new	rat choline kinase D37884 8.0e-5 Pro-rich	nd	
	65	hmc32g10	X88161	244	mouse high S protein M37760 7.0e-5	Cys rish keratin assoc prot X80035 7.4e-7	nd	
D	66	hmc13g09	X88401	142	new	sea urchin sperm flagellar prot A40697 aa 203-233 3.7e-6	nd	
	67	hmc14h07	X88377	144	new	c.elegans T20G5.3 P34576 1.2e-4	nd	
	68	hmc16f07	X88362	125	new	human sperm 75 kD prot S58544 5.5e-4	nd	
	69	hmc17b05	X88353	103	new	human G regulatory prot U02082 aa 216-248 3.5e-5	nd	
	70	hmc18d04	X88333	165	new	yeast Ca-binding prot P34216 aa 321-363 1.3e-4	nd	
	71	hmc11b07	X88420	242	new	P17437 skin secretory protein frog 2.4e-8	nd	
	72	hcm04a04	X88493	228	new	c.elegans T11G6.2 Z69384 3.6e-5	nd	
	73	hmc20a12	X88315	173	new	bovine F10 P00743 5.8e-5	nd	
	74	hmc22g02	X88275	179	new	phytophthora PCSTTK-2 X83423 7.0e-4	nd	

Table 3. *(Continued)*

		Other Homologies						
		Clone ID	GenBank	Size	BLASTN	BLASTX	HC21[a]	Mapping mode[b]
	75	hmc23c11	X88267	125	EST R91742 yp98g05.r1 4.0e-32	E Coli SCRP-27A P26428 aa 66-101 6.6e-7	nd	
	76	hmc26a06	X88227	88	new	klebsiella transposase S38653 aa 129-153 3.3e-1	nd	
	77	hmc29b01	X88187	68	new	human HOX-1D Q00056 1.2e-3	nd	
	78	hmc31h06	X88174	103	new	bovine mucin A60726 4.8e-5	nd	
	79	hmc32h07	X88159	211	new	paramecium G surface protein P13837 7.2e-4	nd	
	80	hmc34c02	X88138	173	new	crambe crambin P01542 4.2e-4	nd	
	81	hmc34e09	X88132	261	new	herpes V nuclear antigen P33485 3.3e-5	nd	
	82	hmc35b02	X88126	205	new	bovine CD5 T cell P19238 2.5e-5	nd	
	83	hmc38a07	X88100	195	EST F11677 nt 202-303 5.8e-46 minus	hamster RNA pol II large subunit P114114 1.0e-4	nd	
	84	hmc42g01	X88065	259	new	rat mineralocorticoid receptor P22199 6.1e-5	nd	
	85	hmc43e01	X88059	193	new	mouse Ig receptor U06431 1.7e-4	nd	
	86	hmc46h07	X88017	167	new	TYMV 69 kd hypothetical protein P10357 6.6e-5	nd	
E	87	hmc09a03	X88440	169	AML1 U19601 nt 862-967 2.5e-34 (-)	HUMAML1B_1 3.3e-17 AML1 region (-)	YES	s
	88	hmc27g11	X88207	114	IFNRAF-1 U05877 nt 1048-1157 2.8e-36 (-)	IFNRAF-1 P38484 aa 313-337 8.5-11 (-)	YES	s
	89	hmc28b09	X88202	302	ETS2 J04102 nt 879-1099 3.1e-83 (-)	ETS2 P15036 aa197-269 1.4e-45 (-)	YES	s
	90	hmc33h01	X88142	111	S. scrofa h2-calponin Z19539 nt 411-510 6.0e-21 (-)	pig calponin h2 Q08094 aa 136-166 2.9e-9 (-)	nd	
	91	hmc36f06	X88109	190	23 kD basic protein X56932 8.6e-41 (-)	human 60S ribosomal protein L13A P40429 3.2e-15 (-)	nd	
F	92	hmc09f06	X88430	261	r(21)Breakpoint M22485 nt 1765-2003 1.2e-91	no homology	YES	y
	93	hmc01a06	X88522	222	EST R78133 yi80f10.r1 1.7e-64	no homology	YES	y,c
	94	hmc04f12	X88481	50	EST R71946 yj84b03.r1 nt216-261 1.3e-6	no homology	nd	
	95	hmc07b04	X88460	154	EST R54917 yj78h05.r1 2.0e-7	no homology	nd	
	96	hmc07d04	X88456	180	EST R25589 yh45b01.r1 6.4e-9	no homology	nd	
	97	hmc07e09	X88454	120	EST R54917 yj78h05.r1 4.2e-12	no homology	nd	
	98	hmc10f07	X88423	120	EST T91946 yd55b05.s1 3.1e-7	no homology	nd	
	99	hmc12e07	X88409	144	EST R96273 yq36e04.r1 1.1e-23	no homology	nd	
	100	hmc13b02	X88408	185	EST N66257 yy68h04.s1 2.6e-36	no homology	nd	
	101	hmc16a05	X88371	87	EST T07990 HIBAA28 nt 34-115 4.6e-27	no homology	YES	c
	102	hmc16b04	X88369	61	EST T07990 HIBAA28 nt 121-174 2.3e-13	no homology	YES	c
	103	hmc16c04	X88367	134	EST H33948 1.1e-22 rat	no homology	YES	h
	104	hmc16g02	X88359	104	EST H35525 1.5e-11 rat	no homology	YES	h
	105	rmc07g11	X88560	90	EST T07990 HIBAA28 nt 176-251 6.0e-20	no homology	YES	c
	106	hmc16g02	X88359	104	EST T89436 ye04a06.s1 9.4e-8	no homology	nd	
	107	hmc17a04	X88355	85	EST R00719 ye74d12.s1 3.7e-20	no homology	nd	
	108	hmc17a06	X88354	79	EST R19767 yg40g05.r1 nt 356-428 5.6e-15	no homology	nd	
	109	hmc17a08	X88351	89	EST R19767 yg40g05.r1 nt 189-273 1.2e-26	no homology	nd	
	110	hmc17c05	X88344	85	EST R19767 yg40g05.r1 nt 274-355 5.6e-26	no homology	nd	
	111	hmc17a07	X88352	51	EST T75374 yc89f07.r1 2.6e-8	no homology	nd	

Table 3. (*Continued*)

		Other Homologies						
		Clone ID	GenBank	Size	BLASTN	BLASTX	HC21[a]	Mapping mode[b]
	112	hmc17f11	X88337	109	EST Z47290 21fi8123 1.7e-34	no homology	YES	c,y,s
	113	hmc19a03	X88323	92	EST R35731 yg67g08.r1 4.9e-28	no homology	nd	
	114	hmc19a09	X88320	146	EST T95686 ye40a04.r1 2.8e-45	no homology	nd	
	115	hmc21d02	X88293	101	HUMPBGDA M95623 7.3e-4 (-)	no homology	nd	
	116	hmc21h06	X88282	92	EST H60051 yr19e01.s1 4.3e-28	no homology	nd	
	117	hmc22b06	X88281	148	EST H52729 yo34a11.s1 2.9e-19	no homology	nd	
	118	hmc22f05	X88277	85	EST T30837 3.1e-14	no homology	nd	
	119	hmc23d08	X88265	81	EST H52729 yo34a11.s1 1.7e-11	no homology	nd	
	120	hmc23e06	X88263	130	EST T81564 yd28d04 1.1e-6	no homology	nd	
	121	hmc25c03	X88242	116	EST N56558 sh1116f 1.7e-36	no homology	nd	
	122	hmc28b03	X88204	125	EST F11677 c-30c01 9.5e-40	no homology	nd	
	123	hmc30a04	X88184	118	L35762 mx1 region nt 14229-14342 5.0e-38	no homology	YES	s
	124	hmc34g04	X88131	169	EST R23544 yg34c12.r1 1.9e-23	no homology	nd	
	125	hmc32a08	X88170	102	EST Z39110 c-10d08 1.8e-24	no homology	nd	
	126	hmc33b02	X88154	63	EST R74138 yj99d04.r1 7.5e-17 224-283	no homology	nd	
	127	hmc33b10	X88153	130	EST R74138 yj99d04.r1 1.0e-27 136-227	no homology	nd	
	128	hmc37d10	X88103	143	EST D31072 7.5e-21	no homology	nd	
	129	hmc39c10	X88090	42	EST Z47315 21fi15D2 4.9e-8	no homology	YES	c,y,s
	130	hmc39f03	X88087	185	EST R20834 yg05c01.r1 6.7e-13	Zea mays trancriptional activator L19495 4.5e-3	nd	
	131	hmc43c10	X88061	60	EST T05687 HFBDE40 6.2e-9	no homology	nd	
	132	hmc44b01	X88052	64	EST T02952 FB17E2 9.3e-5	no homology	nd	
	133	hmc25f07	X88234	44	L35682 clone H8 6-c5 nt 1095-1134 1.9e-8	no homology	YES	s
	134	hmc19b09	X88319	212	L35676 clone H8 2-e7 nt1823-2031 3.2e-77	no homology	YES	s
	135	hmc20h01	X88303	66	L35660 clone H8 6-e2 nt 2013-2074 1.8e-11	no homology	YES	s
	136	hmc30g11	X88176	51	L35659 clone H8 6-h6 nt 1183-1233 4.2e-13	no homology	YES	s
	137	hmc44b11	X88050	102	L35675 clone H8 3-b5 nt 1302-1383 2.3e-29	no homology	YES	s
	138	hmc44c05	X88049	96	L35674 clone H8 4-d4 nt 2225-2301 2.0e-22	no homology	YES	s
	139	hmc44d02	X88047	48	L35679 clone H8 2-d11 nt 45-88 2.0e-9	no homology	YES	s
	140	hmc45c04	X88037	144	EST N47864 yy95d09.s1 1.9e-40	no homology	nd	
	141	rmc06a05	X88547	81	EST L30889 UT1591 9.2e-10	no homology	nd	
	142	rmc06c01	X88549	118	HUMMX1B1 MX1 region L35762 9.7e-38	no homology	YES	s,c
	143	rch06f06	X88550	52	EST T59370 yb57c04.r1 3.4e-4	no homology	nd	
	144	rmc07g07	X88558	63	EST T85467 yd82f03.r1 6.6e-15	no homology	YES	c
G	145	hmc48g01	X88006	45	chr21 trapped & mapped X85357 exSNSD0318	no homology	YES	c,y
	146	hmc48g03	X88005	55	chr21 trapped & mapped R82121 ex18G8	no homology	YES	c,y
	147	hmc47b11	X88014	170	chr21 trapped & mapped R82116 ex17D11	no homology	YES	c,y
	148	hmc43f06	X88058	50	chr21 trapped & mapped R82167 ex8E1	no homology	YES	c,y
	149	hmc27c02	X88215	89	chr21 trapped & mapped R82160 ex7A11	no homology	YES	c,y
	150	hmc24e11	X88251	96	chr21 trapped & mapped R82154 ex5E5	no homology	YES	c,y
	151	hmc20c02	X88314	108	chr21 trapped & mapped R82140 ex3A3	chironomus sp-la K01693 9.0e-5 Pro-rich	YES	c,y

Table 3. (*Continued*)

Other Homologies Clone ID	GenBank	Size	BLASTN	BLASTX	HC21[a]	Mapping mode[b]
152 hmc04a11	X88491	134	chr21 trapped & mapped X85338 exSNS03A06	no homology	YES	c,y
153 hmc03a02	X88498	175	chr21 trapped & mapped X85366 exSNSA03	no homology	YES	c,y
154 hmc01a01	X88524	>305	chr21 trapped & mapped R82158 ex6E9	no homology	YES	c,y
155 hmc01a02	X88523	156	chr21 trapped & mapped R82161/X85344 ex7E6/SNS03D17	no homology	YES	c,y,h

The full-length cDNA sequence and mapping position of the ATP synthase subunit (ATP5O, corresponding to clones A15 and A16) has been reported (Chen et al. 1995a); the mapping position of the human *GABPA* gene (corresponding to clone A1) has been reported (Chrast et al. 1995); the initial characterization and mapping of the human *SIM* gene (corresponding to clones A6–A9) has been published in Chen et al. (1995b).

(*A*) Strong homologies to protein sequences; (*B*) Weak homologies to collagen genes; (C) Weak homologies to proline-rich protein sequences; (*D*) weak homologies to a variety of proteins; (*E*) strong homologies/identities on the "minus" strand (see text); (*F*) strong homologies/identities to ESTs; (*G*) identities to chromosome 21-trapped exons from the study of Lucente et al. (1995).

[a] (nd) Not done.

[b] (c) Chromosome 21 cosmids identified; (y) chromosome 21 YAC identified; (h) mapping by hybrids; (s) mapping by sequence identity to chromosome 21 sequences; (f) mapping by FISH.

chromosome 21 have been published recently (Cheng et al. 1994; Peterson et al. 1994; Xu et al. 1995).

The predicted encoded polypeptides of a total of 49 exons (8.7%) had high homologies to proteins from other species or to related human proteins (Table 3A). This subset of sequences is among the most interesting in the short term because some (albeit hypothetical) predictions can be made about functions, leading potentially to the identification of genes that are candidates for specific phenotypes. These exons include, among others, homologies to the following genes: *Drosophila single-minded, white* locus, and *enhancer of zeste,* rat lanosterol synthase, and megalin, bovine ATP synthase OSCP subunit, yeast PWP2 and protein kinase, *Xenopus* neural cell-adhesion molecule, mouse pericentrin, T-cell invasion and metastasis, and requiem, humam coagulation factor 11 and elastase 2b.

There are several lines of evidence to suggest that the majority of the trapped sequences are fragments of genes: (1) Sequence homology searches identity to many ESTs from various cDNA libraries (Table 3F); (2) a considerable number of homologies to genes from other species have been identified (for some of these, the corresponding full-length cDNA of the human homolog has been cloned; e.g., see Chen et al. 1995a); (3) the GC content of the trapped sequences is more similar to cDNAs than to genomic DNA (for the differences in GC content in the completely sequenced chromosome XI of *Saccharomyces cerevisiae,* see Dujon et al. 1994; for the GC content of the human genome, see Saccone et al. 1993); (4) a number of exons for known chromosome 21 genes have been obtained; and (5) the use of pools of trapped sequences as probes against cDNA libraries identified a substantial number of positive cDNA clones.

The sequencing of the trapped inserts from pAMP10 using oligonucleotide SD2 (see Methods) is directional, that is, from the acceptor toward the donor splice site used, and therefore the homology with known transcripts should always be with the coding strand. In a few instances, however, we have identified significant homologies with the "minus" strand (Table 3E). For example, homologies with $P < 10^{-33}$ were found with regions of the chromosome 21 genes *AML*1, *IFNAR-1,* and *ETS2,* suggesting that there are either transcripts from overlapping genes in opposite directions of that acceptor and donor-like splice sites have been used from the noncoding strand. These possibilities will be tested by isolating and studying any existing corresponding cDNAs.

Exon trapping was applied recently to a 2.5-Mb region of chromosome 21 that has been associated with some features of Down syndrome (Lucente et al. 1995). A total of 102 trapped sequences have been reported and mapped to cosmid clones of the region; the average exon density was 1 in every 25 kb. A total of 13 exons from the present study were identical to those reported in Lucente et al. (1995) (Table 3G; exons for *SIM* and *ERG* genes). Furthermore, both exon trapping and cDNA selection have been applied to 81 cosmids of plate 5 of the chromosome 21-specific LL21NC02-Q cosmid library (Yaspo et al. 1995). After elimination of repetitive elements, a total of 21 apparently different transcription units were identified. Because we did not use plate 5 in our experiments, the cosmids used in the present study and that of Yaspo et al. (1995) were different. Data base searches revealed that our trapped sequences identified portions of four of these transcription units (TU4, 5, 8, 13; Yaspo et al. 1995).

It appears that there are gene-rich and gene-poor regions on 21q. To investigate the gene density, Tassone et al. (1995) used cDNA selection from six cDNA libraries to 16 YACs mapped throughout 21q. They found that the regions 21q22.3 and 21q22.1 are gene-rich as compared with regions 21q11.2, 21q21, and 21q22.2, which yielded very few genes (Yaspo et al. 1995). These results are in agreement with the gene-rich isochores (Saccone et al. 1993), and the distribution of *Not*I restriction sites (Ichikawa et al. 1993) and of CpG islands (Tassone et al. 1992). Our mapping experiments of selected exons, although nonsystematic and complete is in agreement with the results of Tassone et al. (1992); slighty more than half (54%) of the mapped exons localize to the most distal band 21q22.3.

The precise mapping of all the trapped sequences reported here on chromosome 21 and the study of their corresponding cDNAs will enhance our understanding of the gene distribution on 21q and the contribution of this chromosome to human pathologies. In particular, candidate genes involved in Down syndrome and monogenic disorders that map on chromosome 21 can be studied using the clones and nucleotide sequences reported here. We have demonstrated additionally that the exon trapping methodology using DNA material from a single chromosome can be used to isolate portions of the majority of the genes of this chromosome and the completion of the transcription maps. Matching of sequences from these monochromosomal exon trapping experiments with those of the fast progressing EST sequencing programs will immediately provide mapping information on a large number of ESTs and their corresponding clones.

METHODS

Exon Trapping

Genomic DNA cloned in the Lawrence Livermore LL21NCO2-Q chromosome 21-specific cosmid library was used for the exon trapping protocol (instruction manual 18449-017, 1994, GIBCO-BRL). This library, constructed in Lawrist 16 vector, was kindly provided by P. deJong (Soeda et al. 1995). DNA from the 10,368 clones arrayed in 108-microliter plates was spotted onto Hybond membranes (1536 clones per 8 × 12 cm^2 filter). Hybridization of these filters with total mouse DNA or 18S and 28S RNR (probes γ-5.8 and γ-7.3 kindly provided by S. Parimoo, Yale University, New Haven, CT) permitted the recognition of clones with either mouse DNA inserts (1135 clones of 10.9% of the total) or RNR repeats (501 clones or 4.8% of total). The identities of these clones are available on request to haiming@medsun.unige.ch. Pools of 10 cosmid clones from plates Q49 to Q61 and Q63 to Q65 were digested with *Pst*I, and the fragments were subcloned in plasmid pSPL3 (Church et al. 1994). In some experiments, pools of cosmids from entire microliter plates (Q57, Q58, Q60, Q63, Q64, Q21, Q31, Q35, and Q36) were used after digestion with *Eco*RI. Cosmids containing mouse DNA and RNR sequences were excluded from the above experiments. Recombinant plasmids were transfected into cos7 cells using lipofectACE (exon trapping protocol, GIBCO-BRL). After 24 hr, total RNA from cos7 cells was reverse-transcribed and PCR-amplified using primers complementary to pSPL3 sequences (exon trapping protocol, GIBCO-BRL). The reverse transcriptase PCR (RT–PCR) products were subcloned into vector pAMP10 using uracil DNA glycosylase (UDG) cloning (exon trapping protocol, GIBCO-BRL). To eliminate clones that contained false-positive exons attributable to pSPL3 self-splicing (which range from 8%–35% of clones in different experiments), the cloned PCR products were hybridized with oligonucleotides 5′-TAGCAATAGTAGCATTAGTA-3′, 5′-TGCTAAAGCATATGATACAG-3′, 5′-TCATTCTTCAAATCAGTGCA-3′, and 5′-GGATATTCACCATTATCGTT-3′ (which extended from pSPL3 nucleotides 731–750, 1111–1130, 1331–1350, and 3071–3090, respectively) and the positive subclones were eliminated from further analysis.

Nucleotide Sequencing and Data Base Comparisons

The trapped sequences were subjected to nucleotide sequencing with *Taq* polymerase by the dye terminator method using oligonucleotide SD2 5′-GTGAACTGCACTGTGACAAGCTGC-3′ (which is complementary to the 5′ exon provided by the pSPL3 vector) on an AB1373 sequencer. The nucleotide sequences and their predicted translation products in all six reading frames were then used for sequence comparisons against all available nucleotide and protein data bases; the homology search algorithms used were BLASTN and BLASTX (Alschul et al. 1990), and in some cases FASTDB (Brutlag et al. 1990). Data base matches with significance $<10^{-4}$ were considered nonsignificant (unless the test sequence was <50 nucleotides) and the sequences were considered novel. All sequences reported in this papaer have been deposited in the EMBL/GenBank data bases (accession nos. X88001–X88560, X886349–X86351, X83219, X84366, and X83513–X86516).

Genomic Mapping of Trapped Sequences

A subset of the trapped sequences were mapped to chromosome 21 by several methods, including PCR amplification from DNA of YACs, cosmids, or somatic cell hybrids, Southern hybridization, and FISH. Radioactive hybridization probes were prepared from the inserts of selected recombinant pAMP10 plasmids by PCR amplification using oligonucleotides dUSD2 and dUDA4 (exon trapping protocol, GIBCO-BRL). These probes were used for filter hybridization against the chromosome 21 cosmid library LL21NC02-Q, the collection of the YACs from the chromosome 21 YAC contig (Chumakov et al. 1992), and restriction endonuclease-digested genomic DNA from rodent–human somatic cell hybrids containing defined fragments of human chromosome 21 (kindly donated by D. Patterson) (Patterson et al. 1993), from human genomic DNA, and from genomic DNA from yeast clones containing specific YACs. PCR amplification using oligonucleotide primers corresponding to selected trapped sequences was used on template DNA from chromosome 21-specific cosmids, YACs, somatic cell hybrids, and human, mouse, and Chinese hamster DNAs. FISH (Lichter et al. 1988) was performed in a few cases using cosmids positive for certain trapped sequences.

Identification of Corresponding cDNAs

Pools of either 50 or 30 trapped sequences were each used as probes against ~300,000 plaques from humna retina (Nathans et al. 1986) or fetal brain cDNA libraries (Clontech). A proportion of the positive clones were plaque-purified and used as hybridization probes against filters containing the trapped exons.

A number of trapped sequences were used as hybridization probes against cDNA libraries after PCR amplification of the insert. The cDNA library of choice was the normalized infant brain library (Soares et al. 1994). Single clones or pools of five trapped sequences were used for hybridization against the 40,000 clones of this library arrayed in 11 filters (3456 clones per filter). Other cDNA libraries used during the experiments include adult brain, heart, kidney, testis, colon, and 11-week-old human embryo (Clontech; gifts from P. Goodfellow, Cambridge University, UK; D. Kurnit, University of Michigan, Ann Arbor; D. Karagogeos, University of Crete, Herkalion, Greece).

ACKNOWLEDGMENTS

This study was supported by grants from the Swiss Fonds National de Recherche Scientifique 31.33965.92, 31.40500.94, and European Union grant GENE-CT93-0015, and the Geneva University and Hospital funds to S.E.A. We thank Pieter deJong for the chromosome 21-specific cosmid library LL21NC02-Q, David Patterson for the chromosome 21-specific somatic cell hybrids, and Laurent Roux for the cos7 cell line. We also thank Isabelle Moix, Michele Young-Bellido, Sonia Samec, Arnaud Gos, and Jianying Su for their contributions in some experiments and Amos Bairoch for help in data base searches. The sequence data described in this paper have been submitted to the EMBL/GenBank data library under accession numbers X88001–X88560, X86349–X86351, X83219, X84366, and X83513–X86516.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

Altschul, S.F., W. Gish, W. Miller, E.W. Myers, and D.J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215:** 403–410.

Antonarakis, S.E. 1993. Human chromosome 21: Genome mapping and exploration, circa 1993. *Trends Genet.* **9:** 142–148.

Antonarakis, S.E., D. Patterson, C. Van Broeckhoven, N. Shimizu, K. Gardiner, J. Delabar, J. Korenberg, and R. Reeves. 1995. Report of the committee on the genetic constitiution of chromosome 21. In *Human gene mapping 1994* (ed. A.J. Cuttichia), pp. 732–766. Johns Hopkins University Press, Baltimore, MD.

Berry, G.T., J.J. Mallee, H.M. Kwon, J.S. Rim, W.R. Mulla, M. Muenke, and N.B. Spinner. 1995. The human osmoregulatory Na$^+$/myo-inositol cotransporter gene (SLC5A3): Molecular cloning and localization to chromosome 21. *Genomics* **25:** 507–513.

Brennan, M.B. and U. Hochgeschwander. 1995. So many needles, so much hay. *Hum. Mol. Genet.* **4:** 153–156.

Brutlag, D.L., J.P. Dautricourt, S. Maulik, and J. Relph. 1990. Improved sensitivity of biological sequence database searches. *Comp. Appl. Biosci.* **6:** 237–245.

Buckler, A.J., D.D. Chang, S.L. Graw, J.D. Brook, D.A. Haber, P.A. Sharp, and D.E. Housman. 1991. Exon amplification: A strategy to isolate mammalian genes based on RNA splicing. *Proc. Natl. Acad. Sci.* **88:** 4005–4009.

Chen, H., M.A. Morris, C. Rossier, J.L. Blouin, and S.E. Antonarakis. 1995a. Cloning of the cDNA for the human ATP synthase subunit (ATP5O) by exon trapping and mapping to chromosome 21 *Genomics* **28:** 470–476.

Chen, H.M., R. Chrast, C. Rossier, A. Gos, S.E. Antonarakis, J. Kudoh, A. Yamaki, N. Shindoh, H. Maeda, S. Minoshima, and N. Shimizu. 1995b. Single-minded and Down syndrome? *Nature Genet.* **10:** 9–10.

Cheng, J.F., V. Boyartchuk, and Y. Zhu. 1994. Isolation and mapping of human chromosome 21 cDNA: Progress in constructing a chromosome 21 expression map. *Genomics* **23:** 75–84.

Chrast, R., H.M. Chen, M.A. Morris, and S.E. Antonarakis. 1995. Mapping of the human transcription factor GABPA gene to chromosome 21. *Genomics* **28:** 119–122.

Chumakov, I., P. Rigault, S. Guillou, P. Ougen, A. Billaut, G. Guasconi, P. Gervy, I. LeGall, P. Soularue, L. Grinas, L. Bougueleret, C. Bellanne-Chantelot, B. Lacroix, E. Barillot, P. Gesnouin, S. Pook, G. Vaysseix, G. Frelat, A. Schmitz, J.L. Sambucy, A. Bosch, X. Estivill, J. Weissenbach, A. Vignal, H. Riethman, D. Cox, D. Patterson, K. Gardiner, M. Hattori, Y. Sakaki, H. Ichikawa, M. Ohki, D. Le Paslier, R. Heilig, S.E. Antonarakis, and D. Cohen. 1992. A continuum of overlapping clones spanning the entire chromosome 21q. *Nature* **359:** 380–386.

Church, D.M., L.T. Banks, A.C. Rogers, S.L. Graw, D.E. Housman, J.F. Gusella, and A.J. Buckler. 1993. Identification of human chromosome 9 specific genes using exon amplification. *Hum. Mol. Genet.* **2:** 1915–1920.

Church, D.M., C.J. Stotler, J.L. Rutter, J.R. Murrell, J.A. Trofatter, and A.J. Buckler. 1994. Isolation of genes from complex sources of mammalian genomic DNA using exon amplification. *Nature Genet.* **6:** 98–105.

Collins, F. and D. Galas. 1993. A new five-year plan for the U.S. human genome project. *Science* **262:** 43–46.

Dujon, B., D. Alexandraki, B. André, W. Ansorge, V. Baladron, J.P. Ballesta, A. Banrevi, P.A. Bolle, M. Bolotin-Fukuhara, P. Bossier, G. Bou, J. Boyer, M.J. Buitrago, G. Chéret, L. Colleaux, B. Daignan-Fornier, F. del Rey, C. Dion, H. Domdey, A. Düsterhöft, S. Düsterhus, K.-D. Entian, H. Erfle, P.F. Esteban, H. Feldmann, L. Fernandes, G.M. Fobo, C. Fritz, H. Fukuhara, C. Gabel, L. Gaillon, J.M. Carcia-Cantalejo, J.J. Garcia-Ramirez, M.E. Gent, M. Ghazvini, A. Goffeau, A. Gonzalez, D. Grothues, P. Guerreiro, J. Hegemann, N. Hewitt, F. Hilger, C.P. Hollenberg, O. Horaitis, K.J. Inge, A. Jacquier, C.M. James, J.C. Jauniaux, A. Jimenez, H. Keuchel, L. Kirchrath, K. Kleine, P. Kötter, P. Legrain, S. Liebl, E.J. Louis, A. Maia e Silva, C. Mark, A.-L. Monnier, D. Möstl, S. Müller, B. Obermaier, S.G. Oliver, C. Pallier, S. Pascolo, F. Pfeiffer, P. Philippsen, R.J. Planta, F.M. Pohl, T.M. Pohl, R. Pöhlmann, D. Portetelle, B. Purnelle, V. Puzos, M. Ramezani Rad, S.W. Rasmussen, M. Remacha, J.L. Revueita, G.-F. Richard, M. Rieger, C. Rodrigues-Pousada, M. Rose, T. Rupp, M.A. Santos, C. Schwager, C. Sensen, J. Skala, H. Soares, F. Sor, J. Stegemann, H. Tettelin, A. Thierry, M. Tzermia, L.A.

Urrestarazu, L. van Dyck, J.C. van Vliet-Reedijk, M. Valens, M. Vandenbol, C. Vilela, S. Vissers, D. von Wettstein, H. Voss, S. Wiemann, G. Xu, J. Zimmermann, M. Haasemann, I. Becker, and H.W. Mewes. 1994. Complete DNA sequence of yeast chromosome XI. *Nature* **369:** 371–378.

Elson, A., D. Levanon, M. Brandeis, N. Dafni, Y. Bernstein, E. Danciger, and Y. Groner. 1990. The structure of the human liver-type phosphofructokinase. *Genomics* **7:** 47–56.

Fields, C., M.D. Adams, O. White, and J.C. Venter. 1994. How many genes in the human genome? *Nature Genet.* **7:** 345–346.

Ichikawa, H., F. Hasada, Y. Arai, K. Shimizu, M. Ohiva, and M. Ohki. 1993. A NotI restriction map of the entire long arm of human chromosome 21. *Nature Genet.* **4:** 361–366.

Krizman, D.B. and S.M. Berget. 1993. Efficient selection of 3′-terminal exons from vertebrate DNA. *Nucleic Acids Res.* **21:** 5198–5202.

Leung, T., B.E. How, E. Manser, and L. Lim. 1994. Cerebellar beta 2-chimaerin, a GTPase-activating protein for p21 ras-related rac is specifically expressed in granule cells and has a unique N-terminal SH2 domain. *J. Biol. Chem.* **269:** 12888–12892.

Lichter, P., T. Cremer, J. Borden, L. Manuelidis, and D.C. Ward. 1988. Delineation of individual human chromosomes in metaphase and interphase cells by in situ suppression hybridization using recombinant DNA libraries. *Hum. Genet.* **80:** 224–234.

Lovett, M., J. Kere, and L.M. Hinton. 1991. Direct selection: A method for the isolation of cDNAs encoded by large genomic regions. *Proc. Natl. Acad. Sci.* **88:** 9628–9632.

Lucente, D., H.M. Chen, D. Shea, S.N. Samec, R. Chrast, C. Rossier, A. Buckler, S.E. Antonarakis, and M.K. McCormick. 1995. Localization of 102 exons to a 2.5 Mb region of chromosome 21 involved in Down syndrome. *Hum. Mol. Genet.* **4:** 1305–1311.

McInnis, M.G., A. Chakravarti, J. Blaschak, M.B. Petersen, V. Sharma, D. Avramopoulos, J.L. Blouin, U. Koenig, C. Brahe, T. Cox Matise, A.C. Warrren, C.C. Talbot, Jr., C. Van Broeckhoven, M. Litt, and S.E. Antonarakis. 1993. A linkage map of human chromosome 21: 43 markers at average interval of 2.5 cM. *Genomics* **16:** 562–571.

Murai, T., A. Kakizuka, T. Takumi, H. Ohkubo, and S. Nakanishi. 1989. Molecular cloning and sequence analysis of human genomic DNA encoding a novel membrane protein which exhibits a slowly activating potassium channel activity. *Biochem. Biophys. Res. Commun.* **161:** 176–181.

Nathans, J., D. Thomas, and D.S. Hogness. 1986. Molecular genetics of human color vision: The genes encoding blue, green, and red pigments. *Science* **232:** 193–202.

Nizetic, D., L. Gellen, R.M.J. Hamvas, R. Mott, A. Grigoriev, R. Vatcheva, G. Zehetner, M.L. Yaspo, A. Dutriaux, C. Lopes, J.M. Delabar, C. Van Broeckhoven, M.C. Potier, and H. Lehrach. 1994. An integrated YAC-overlap and "cosmid-pocket" map of the human chromosome 21. *Hum. Mol. Genet.* **3:** 759–770.

Parimoo, S., S.R. Patanjali, H. Shukla, D.D. Chaplin, and S.M. Weissman. 1991. cDNA selection: Efficient PCR approach for the selection of cDNAs encoded in large chromosomal DNA fragments. *Proc. Natl. Acad. Sci.* **88:** 9623–9627.

Patterson, D., Z. Rahmani, D. Donaldson, K. Gardiner, and C. Jones. 1993. Physical mapping of chromosome 21. *Progr. Clin. Biol. Res.* **384:** 33–50.

Peterson, A., N. Patil, C. Robbins, L. Wang, D.R. Cox, and R.M. Myers. 1994. A transcript map of the Down syndrome critical region on chromosome 21. *Hum. Mol. Genet.* **3:** 1735–1742.

Saccone, S., A. De Sario, J. Wiegant, A.K. Raap, G. Della Valle, and G. Bernardi. 1993. Correlations between isochores and chromosomal bands in the human genome. *Proc. Natl. Acad. Sci.* **90:** 11929–11933.

Soares, M.B., M.D. Bonaldo, P. Jelene, L. Su, L. Lawton, and A. Efstratiadis. 1994. Construction and characterization of a normalized cDNA library. *Proc. Natl. Acad. Sci.* **91:** 9228–9232.

Soeda, E., D.X. Hou, K. Osoegawa, Y. Atshuchi, T. Yamagata, T. Shimokawa, H. Kishida, E. Soeda, S. Okano, I. Chumakov, D. Cohen, M. Raff, K. Gardiner, S.L. Graw, D. Patterson, P.D. Jong, L.K. Ashworth, T. Slezak, and A.V. Carrano. 1995. Cosmid assembly and anchoring to human chromosome 21. *Genomics* **25:** 73–84.

Tassone, F., S. Cheng, and K. Gardiner. 1992. Analysis of chromosome 21 yeast artificial chromosome (YAC) clones. *Am. J. Hum. Genet.* **51:** 1251–1264.

Tassone, F., H. Xu, H. Burkin, S. Weissman, and K. Gardiner. 1995. cDNA selection from 10 Mb of chromosome 21 DNA: Efficiency in transcriptional mapping and reflections of the genome organization. *Hum. Mol. Genet.* **4:** 1509–1518.

Watanbe, H., J.I. Sawada, K.I. Yano, K. Yamagucchi, M. Goto, and H. Handa. 1993. cDNA cloning of transcription factor E4TF1 subunits with ETS and Notch motifs. *Mol. Cell. Biol.* **13:** 1385–1391.

Xu, H., H. Wei, F. Tassone, S. Graw, K. Gardiner, and S.M. Weissman. 1995. A search for genes from the dark band regions of human chromosome 21. *Genomics* **27:** 1–8.

Yaspo, M.-L., L. Gellen, R. Mott, B. Korn, D. Nizetic, A. Poustka, and H. Lehrach. 1995. Model for a transcript map of human chromosome 21: Isolation of new coding sequences from exon and enriched cDNA libraries. *Hum. Mol. Genet.* **4:** 1291–1304.

Received May 20, 1996; accepted June 21, 1996.