Human Molecular Genetics, 1995, Vol. 4, No. 8 1305–1311

Localization of 102 exons to a 2.5 Mb region involved in Down syndrome

D.Lucente, H.M.Chen[1], D.Shea, S.N.Samec[1], M.Rutter, R.Chrast[1], C.Rossier[1], A.Buckler, S.E.Antonarakis[1] and M.K.McCormick*

Molecular Neurogenetics Unit, Department of Neurology, Massachusetts General Hospital and Harvard Medical School, Charlestown, MA 02140, USA and [1]Genes 'R' Us Laboratory of Human Molecular Genetics, Department of Genetics and Microbiology, University of Geneva Medical School and Division of Medical Genetics, University Cantonal Hospital of Geneva, Switzerland

Received April 10, 1995; Revised and Accepted June 19, 1995

Exon amplification has been applied to a 2.5 Mb region of chromosome 21 that has been associated with some features of Down syndrome (DS). Identification of the majority of genes from this region will facilitate the correlation of the over-expression of particular genes with specific phenotypes of DS. Over 100 gene fragments have been isolated from this 2.5 Mb segment. The exons have been characterized by sequence analysis, comparison with public databases and expansion to cDNA clones. Localization of the exons to chromosome 21 has been determined by hybridization to genomic Southern blots and to YAC and cosmid clones representing the region. This has resulted in a higher resolution physical map with a marker approximately every 25 kb. This integrated physical and transcript map will be valuable for fine mapping of DNA from individuals with partial aneuploidy of chromosome 21 as well as for assessing and ultimately generating a complete gene map of this segment of the genome.

INTRODUCTION

Since the association of Down syndrome (DS) with trisomy for human chromosome 21 (1,2), the study of this common birth defect has progressed from cytogenetic analyses of patient chromosomes (3) through the molecular characterization of gene and sequence dosage in individuals with partial trisomy 21 (4–9). The resulting genotype-phenotype correlations that have emerged define subregions of the chromosome that likely contain genes contributing to specific features of DS (5,10).

The current correlation of DS phenotypes with sub-regions of chromosome 21 has relied primarily on the analyses of sequence copy number in individuals with partial trisomy 21. In one model, the genotype-phenotype definitions are based on the smallest region of overlap (SRO) present in triplicate in different individuals with partial trisomy 21 (5). The features of the syndrome present in common in these individuals are assigned to genes in the SRO. In an alternative model, if more than one gene on chromosome 21 contributes to a DS phenotype, then the minimal molecular overlap in the DS individuals exhibiting the phenotype may erroneously identify the region believed to contain the gene(s) (10). Those features of DS that are likely to be influenced by more than one gene on chromosome 21 are identified by comparing the penetrance and expressivity of the trait in the DS and the unaffected populations. In both models, the region around locus D21S55, which is between loci CBR and ERG, is almost always implicated in contributing to DS phenotypes and has only been excluded from contributing to one phenotype (10). Therefore, we have chosen to concentrate our initial gene isolation efforts on this interval, while also recognizing that additional regions of chromosome 21 likely contain genes of significance to DS. Although defining these regions is a complex issue, an obvious next step toward eventual understanding of the etiology of DS is to begin isolating the genes from these segments of the chromosome. It is precisely because of this complexity, and the number of genes that may have to be isolated and characterized, that DS is an excellent choice for the application of large scale, comprehensive gene isolation efforts.

Many techniques exist for isolating expressed sequences, including utilizing CpG islands or evolutionary sequence conservation as indicators of the presence of genes, hybridization of genomic fragments to cDNA libraries, direct cDNA selection and exon amplification. The last two methods are most amenable to gene isolation from megabase genomic regions. The successful application of direct cDNA selection for identifying genes from chromosome 21 has recently been reported (11–13). We have utilized exon amplification to isolate over 100 gene fragments from a 2.5 Mb region of chromosome 21 between loci CBR and ERG, which is present in three copies in common in many DS individuals with partial trisomy 21 (4,5,7,10). Localization of these gene fragments to a physical map of the region has imparted higher resolution to the physical map and contributes to integration with a transcript map.

RESULTS

Exon isolation and integration with the physical map

A group of 12 YAC clones previously localized in the ~2.5 Mb distance between loci CBR and ERG was used to identify 394 corresponding cosmids from a flow sorted chromosome 21 cosmid library, LL21NC02, constructed at Lawrence Livermore National Laboratory (Fig. 1). Sixteen markers previously localized to this genomic segment were assayed in this subset

*To whom correspondence should be addressed

of cosmids. Markers that were not represented in this subset were used to identify additional clones from the cosmid library by PCR. Gene fragments were then isolated from pools of these cosmids by the method of exon amplification (14). Background was eliminated by hybridization of trapped background products to colony grids of all the trapped products. Redundant exon clones relative to those exons in the first row of twelve were detected by hybridization of individual exons to colony grids of all the exon clones resulting in 404 potentially unique gene fragments. Complete sequence has been obtained for 115 exons and 102 exons have been localized on the physical map.

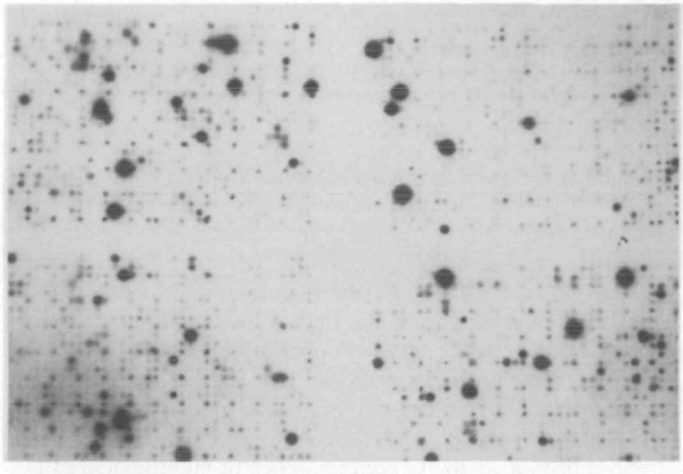

Figure 1. Identification of cosmid clones corresponding to YACs. Cosmid colony grids of a portion of the chromosome 21 cosmid library are shown. They have been hybridized to pooled *Alu*-PCR products generated individually from four YACs. Positive clones were transferred to 96-well dishes and re-screened with each of the YAC *Alu*-PCR products individually. Cosmids that re-screened positive were pooled for subsequent isolation of exons.

The chromosome location of exons was confirmed in two ways. Forty one exons were mapped to chromosome 21 by hybridization to Southern blots containing genomic DNA from human and rodent cell lines and somatic cell hybrids (Fig. 2A). A total of 102 exons, or 1 exon every ~25 kb on average, have been localized to the physical map of the 2.5 Mb between CBR and ERG by hybridization to YAC and/or cosmid colony grids (Fig. 2B). Overlapping clones will contain some exons in common. Cosmid clones within and between cosmid bins were assembled based on the presence or absence of the exons. Hybridization of the exons to the cosmids resulted in organizing 276 cosmid clones into 34 contigs covering ~2.2 Mb and leaving ~300 kb distributed in 33 gaps. This is an average gap size of ~10 kb, with a range from <10 kb to ~100 kb. The integrated exon, cosmid and YAC physical map is illustrated in Figure 3. The cosmid clones associated with each contig are listed in Table 1 and can also be obtained from the anonymous ftp site at neuron.mgh.harvard.edu by logging in as anonymous, giving your email as the password and 'cd' to the directory /pub. The file is named dscosex. The ftp site may also be accessed via a World Wide Web (WWW) browser through our chromosome 9q34 Home Page at URL http://neuron.mgh.harvard.edu:8000.

Comparison with other cosmid contigs

The cosmid contigs generated by exon content were independently identified and then compared to cosmid contigs, A

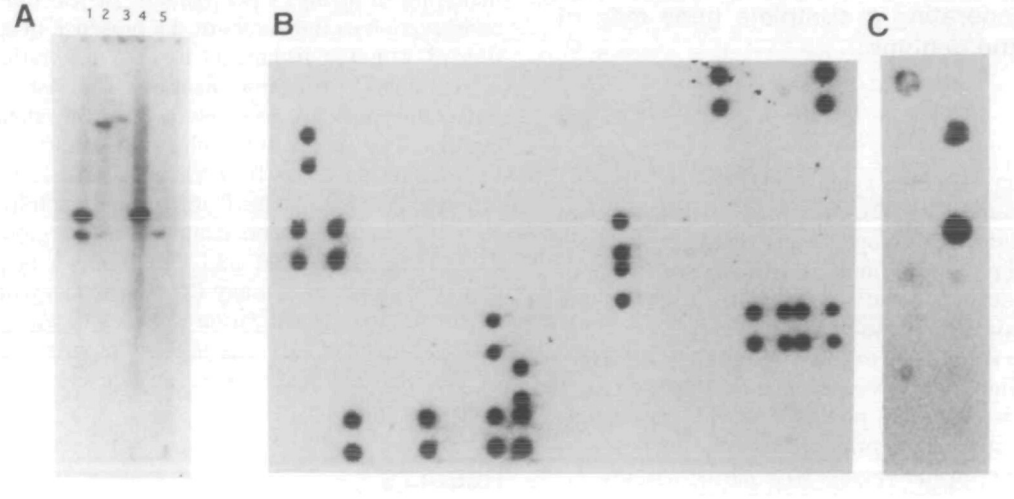

Figure 2. Localization of exons to chromosome 21. (**A**) Genomic Southern blot hybridized with exon 4E2. Digested genomic DNA, isolated from cell lines, was separated on 1.0% agarose gels, transferred to nylon membranes and hybridized with gel purified exon clones following PCR amplification of the exon from the cloning vector with T3 and T7 primers. Lanes: 1, mouse-human hybrid containing human chromosome 21; 2, hamster-mouse hybrid containing mouse chromosome 16; 3, hamster; 4, human; 5, mouse. The common band in lanes 1 and 4 indicate this exon localizes to chromosome 21. Similarly, a common band in lanes 2 and 5 indicate cross hybridization to mouse genomic DNA and localization to mouse chromosome 16. This sequence is also conserved in hamster genomic DNA as indicated by the bands in lanes 2 and 3. (**B**) Cosmid colony grid hybridized with exon 6E3. Colony grids of the 394 cosmids identified by hybridization with YAC *Alu*-PCR products were prepared using a Biomek 1000 robot. Clones were stamped in duplicate in a 4×2 array. Cosmids containing the exon are indicated by the duplicate signals. Exon 6E3 is present in 17 cosmids. (C) YAC colony grids hybridized with exon 6E3. Colony grids from a subset of YACs localized between CBR and ERG-2, made as described (26) are shown. Only those YACs used to identify cosmids are included in Figure 3. Column 1, top to bottom, 751B9,750F7, A222A12, 259H11, C3C1, 413C1, 179B7; Column 2, top to bottom, 152F7, B139B9, 285E6, 336G11, 175G9, 549A9, 743A4. Exon 6E3 is present in YACs 751B9, B139B9 and 336G11.

through Y, in this region described by Patil *et al.* (12). The locations of those contigs relative to this map are shown in Figure 3. In general, the majority of contigs are consistent between the two maps. However, exon mapping has provided evidence for clarifying some previously ambiguous contig orders and connecting cosmid contigs. For example, previously described and unordered contigs D and G are connected by our bins 11 and 12, suggesting the order E-D-G when the STS content of cosmids in E and G contigs are also considered. Similarly, our bins 47–50 and 53 connect previously described contigs L and N, and N and O, respectively. Finally, contigs V and X are contiguous based on an exon (8B11) in common between cosmid 23H5 (contig V) and 9C9 (contig X). The order VX, as opposed to XV, is suggested based on the presence of ERG exons in a larger contig containing VX and the known orientation of the ERG gene (15). Some minor discrepancies were also uncovered as a result of comparing the two cosmid maps. For example, some cosmids were described as disjoint (12) that we found to be overlapping based on exon content, including cosmids 3E2 and 45A8; 19F7, 28B8 and 34H12; 56H1 and 41B8; 39D1 and 47D9; 56H1 and 39D1.

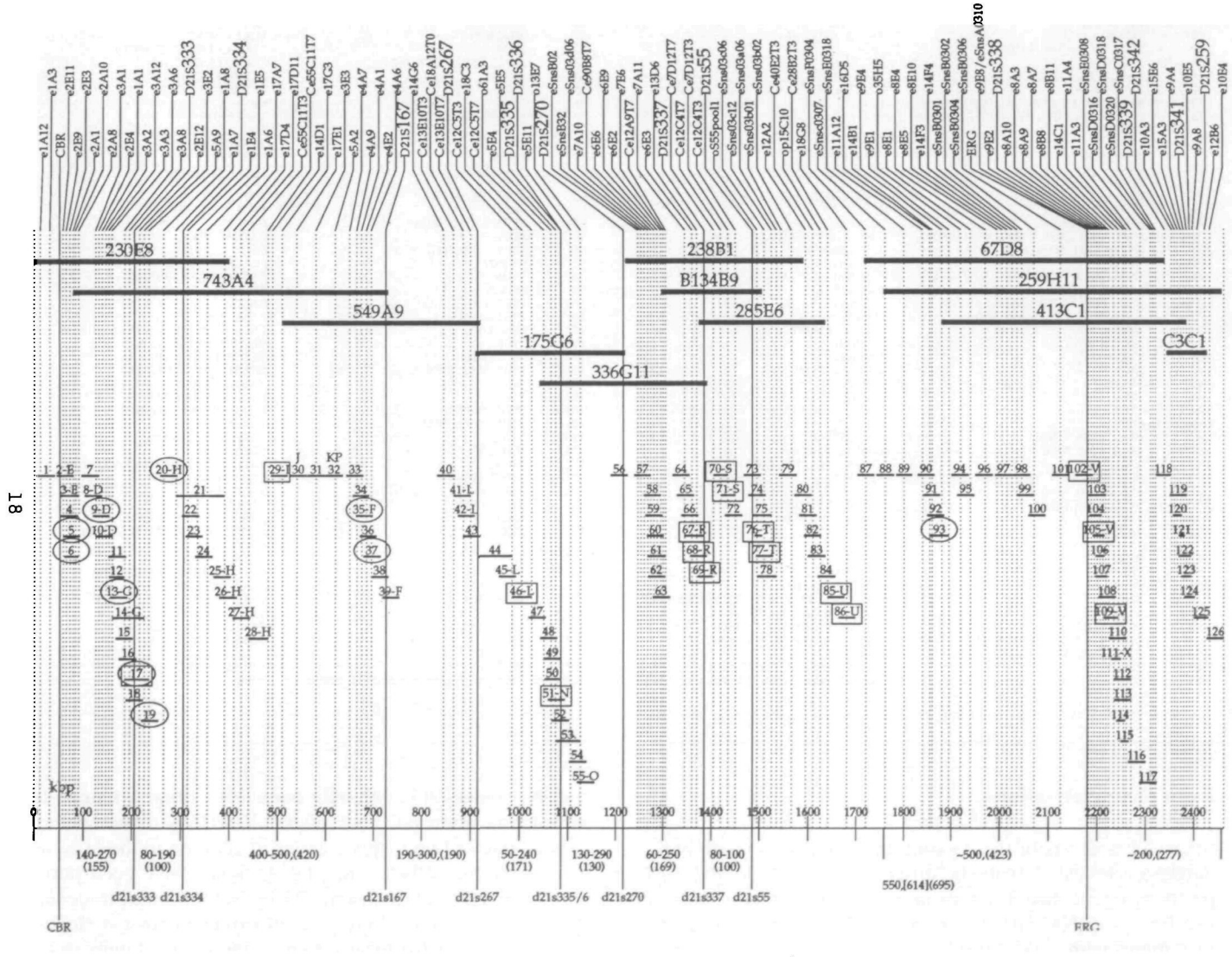

Figure 3. Integrated exon, cosmid and YAC map of the CBR-ERG region of chromosome 21. Markers are listed across the top of the figure. They are: exons (beginning with 'e'), known genes (capitalized), cosmid end clones (beginning with 'Ce') and STS. Markers beginning with 'o' indicate detection of a cosmid clone by hybridization with a whole cosmid probe. There is a corresponding line from each marker through the YAC and/or cosmid clones containing it to the kilobase scale at the bottom of the figure. YACs are depicted below the markers by the bold, labeled lines. Cosmid bins are shown staggered and numbered 1–126 below the YACs. Bins consist of individual or multiple cosmids as detailed in Table 1. Individual cosmids are arranged as overlapping clones based primarily on exon content. The cosmid location of the 10 cDNA clones that have been expanded from exons is indicated by a circle. Additional detail regarding markers present in specific cosmid clones is given in Table 1. An attempt has been made to coordinate previously published mapping information about this region with this exon map. Therefore, the cosmids are annotated with letters to reflect those clones that have also been reported (12). Estimates of fragment sizes between various markers based on pulsed field gel electrophoresis and restriction fingerprinting of YACs (15,19,25) are given below the bottom scale along with the estimated size of the same fragment from this map in parentheses. The cosmid location of cDNA clones reported by other groups (11,13) is indicated by a square if the cosmid was also present in this collection of clones.

Table 1. Cosmid clones and associated exons

Cosmid ID #	Cosmid Names	Exons and Markers Present
1	13B6	1A12, 1A3
2	41H4	CBR, 2E11, 2E9, 2E3, 2A1
3	43F1	CBR, 2E11, 2E9, 2E3, 2A1, 2A10, 2A8
4	44E9, 71E4	2E11, 2E9, 2E3, 2A1, 2A10
5	1D1, 1G5	2E11, 2E9, 2E3, 2A1, 2A10, 2A8
6	13F6, 60D5	2E9, 2E3, 2A1, 2A10
7	79C10	3A1
8	13B3, 17E6, 88H1	3A1, 2E4
9	3E2	3A1, 2E4, 1A1, 3A2, 3A12
10	2H2, 34A8, 37B11, 37F11 45A8	3A1, 2E4, 1A1, 3A2, 3A12, 3A3
11	32E5,47E6, 56B8	3A12, 3A3, 3A6
12	13A4, 75A9, 82G12, 95B5, 105G10	3A3
13	3B11, 45C1, 74E9	3A3, 3A6
14	68E4	3A3, 3E2
15	7F2, 64F7, 76D5, 101F9, 101G9, 101G10, 103B8, 103C11, 103D12, 106D2, 107G12	3A6, 3A8
16	71C8, 103H11, 107H12	3A6, 3A8, D21S333
17	1D9, 56H3	3A6, 3A8, 3E2, 5A9
18	74F10	3A8, D21S333, 2E12
19	52D3, 54G4, 65A6, 102D9	2E12
20	28C6, 85E5	3E2, 5A9
21	17B6, 62F5, 70G2, 70H2	1A8, 1A7
22	34F2, 105A4	D21S334, 1E4, 1E5
23	8A9, 23F9, 35H2, 83C4, 105B2	1E4, 1E5
24	19E7, 55G6, 55H5, 88B10.105E1, 105E4, 109H1	1E5, 1A6
25	41B8	17A7
26	47D9, 56H1	17A7, 17D4
27	39D1	17D4, 17D11, Ce55C11T3
28	55C11	
29	5E7	14D1, 17G3
30	46G4	17E1
31	95E5, 106B12	3E3
32	19F7, 19H3, 28B8, 34H12, 43F4, 85D11	5A2
33	49D2, 98E11, 109B11	4A7, 4A9, 4A1
34	49G1	4A9, 4A1, 4E2
35	14D6, 15A3, 41G5	4A9, 4A1, 4E2, 4A6
36	66A4	4A1, 4E2, 4A6
37	32A1, 38E12, 47E12, 68D10, 68E10	4E2, 4A6
38	22D4, 71B7, 72D2, 83G9, 86D7	D21S167
39	44D5, 99B5	14G6, Ce13E10T3
40	24E1, 64E8	Ce13E10T3
41	13E10	Ce18A12
42	36C1	Ce18A12, Ce13E10T7
43	18A12	D21S267
44	19G9	Ce12C5T3
45	12C5	18C3
46	50D7	5E5
47	61A3, 62H6, 108F3	Ce12C5T7
48	27C9, 30A8, 83E6	Co61A3, 5E4, 5E5
49	57E1, 83E4	Co61A3, 5E4, 5E5, D21S335/336
50	9E3, 17F2, 24G1, 28A9, 50F8, 51H5, 56H8	5E4, 5E5
51	4C3, 24C10, 34E6, 89C2	5E4, 5E5, D21S335/336
52	51D6, 99B9, 102A3	5E5, D21S335/336
53	13E7	5E5, D21S335/336, 5E11
54	1A7, 25H8	5E11, Co13E7
55	43G2	Co13E7
56	2E5	D21S270
57	90B8	SNSB02, SNSB32
58	4F4	SNS03D06, 7A10, 6E6, 6E9, 6E2, Ce90B8T7
59	11A1	7A10, 6E6, 6E9, 6E2, Ce90B8T7
60	69A6, 69B7	7A10, 6E6, 6E9, 6E2, Ce90B8T7, 7E6
61	11B1, 12A9, 31B10, 31D8, 35F12, 67E2, 67F2, 67G2, 80G5	7A10, 6E6, 6E9, 6E2, Ce90B8T7
62	109C5	6E6, 6E9, 6E2, Ce90B8T7, 7A6
63	55A8	6E6, 6E9, 6E2, Ce90B8T7, 7A6, Ce12A9T7
64	43B3	7A11
65	23H3	7A11, 6E3
66	22E9, 22G8, 27C6, 27E2, 43A10, 59B6, 67E11, 75G4, 82C5, 82D10, 86F11, 86G10, 103H8, 103H10	7A11, 6E3, 13D6
67	40D7	7A11, 6E3, 13D6, D21S337
68	45B11	6E3, 13D6, D21S337
69	5A8	D21S337
70	7D12	Ce12C4T7
71	12C4	Ce7D12T3
72	102C2	Ce7D12T3, Ce12C4T3
73	85E11, 95E12, 95F12	D21S55, Co43D4
74	43D4	D21S55, SNS03C06
75	93C11	D21S55, Co43D4 SNS03C06, SNS03C12
76	51G9, 40B10	Co43D4
77	12G3	Co43D4 SNS03C06, SNS03C12
78	11D8	Co43D4 SNS03C06, SNS03C12, SNS03A06
79	94H7	SNS03B01, SNS03B02
80	103G8	12A2
81	5G4, 30F1	12A2, Ce40E2T3
82	30G1	12A2, Ce40E2T3, Co15C10
83	22D12, 56D2, 72C7, 76B10	Ce40E2T3, Co15C10
84	15C10	Ce28B2T3, Co15C10
85	28B2	Ce28B2T3
86	35A2	18G8
87	88A10, 90F6	SNSF0304
88	74D7	SNSC0307
89	83F10	SNS03B01
90	6E12, 16B11, 21G2	11A12, 16D5, 14B1
91	108G5	16D5, 14B1
92	3D11, 76G11	16D5, 14B1, 9E4
93	98H10	16D5, 14B1, 9E4, 9E1
94	35H5, 35H11, 58H5	Co35H5
95	98B9	Co35H5, 8E1
96	22D1, 63E6, 70H1, 73D10	8E4
97	97D10	8E5, 8E10, 14F3, 14F4
98	73H5	SNSB0301, SNSB0302
99	69F12	SNSB0301, SNSB0302, SNSB0304
100	18E3, 24B12, 73G2, 79G4	SNSB0304
101	29A10, 29G11, 66G2, 70E3, 80B4	SNSB0306
102	8B6, 45F5	ERG, 9E8/SNSA03, SNSA0310, 9E2
103	97F8	9E2, D21S338, 8A10, 8A9
104	7G5, 35B11	9E2, D21S338, 8A10, 8A3
105	33B8, 56F2, 57E5, 87G12	D21S338, 8A10, 8A3, 8A9
106	51B5	8A10, 8A3, 8A9
107	51E4	8A10, 8A3, 8A9, 8A7
108	76G5, 84B9	8A3, 8A9, 8A7, 8B8, 8B11
109	23H5, 29B2, 29H3	8A7, 8B8, 8B11
110	54G7	8B11, 14C1, 11A4
111	9C9, 29C4	8B11, 14C1
112	83C7	8B11, 14C1, 11A4, 11A3, SNSE0308
113	102E3, 106E8	14C1, 11A4, 11A3
114	48E7, 49A11	14C1, 11A4
115	11H8, 75H2	11A4, 11A3
116	68G2	SNSE0308, SNSD0316
117	75E2	SNSD0316, SNSD0318, SNSD0320, SNSD0317
118	36F10	D21S339
119	59F11	D21S342, 15A3, 9A4, D21S341
120	78B12	15A3, 15E6, 10A3
121	44G4,53E8	15A3, 9A4
122	104F3	15A3, 9A4, D21S341, 10E5
123	25F8, 71A2	9A4, D21S341, 10E5
124	78D12	D21S341, 10E5, 9A8
125	26B9	D21S259, 12B6
126	32G1, 35F3, 44E8, 59C6, 78H1, 82D5	10E4

Exon characterization

Sequence analysis of these exons and comparison with public databases revealed that six different exons from the ERG proto-oncogene had been isolated. Several other significant similarities [$p(N) > 10^{-4}$] were found. The strongest similarity was found with exon 17D11 and the *Drosophila* gene single-minded (sim), which is a transcriptional regulator of genes controlling development of some cells in the *Drosophila* central nervous system (16,17). The results of the database analyses are listed in Table 2.

The authenticity of the isolated exons as parts of genes is supported by expansion of the exons to cDNA clones and/or identification of a mRNA species by hybridization of the exon or cDNA fragment to northern blots. To date, 16 exons have identified 10 different cDNA clones by hybridization to one or more cDNA libraries. Nine other exons have not identified a cDNA clone, although only one cDNA library (human fetal brain) was screened. Four of the 10 cDNA clones obtained contain one or two mapped exons in addition to the exon used to identify the cDNA. The cDNA clones have been purified and localized to chromosome 21 by hybridization to genomic Southern blots and colony grids of exon and cosmid clones as described above for localization of the exons. Figure 4 is an example of a cDNA hybridization to cosmid and exon colony grids. A representative cosmid to which each of the 10 cDNA clones has been localized is indicated in Figure 3 by a circle. Three exons and three cDNA fragments expanded from an exon have been hybridized to northern blots to determine expression pattern and mRNA size. Figure 5 is an example of a northern blot hybridized with a cDNA fragment that was expanded from exon 3E2. This cDNA detects a predominant ~8 kb mRNA species that appears to be expressed at varying levels in all adult tissues assessed.

Table 2. Results of sequence database searches

#	EXON	MAP kbp	BLAST N highest scoring match	P(N)	BLAST X highest scoring match	P(N)
1	2E11	57	yb07e10.r1 Homo sapiens cDNA clone	1.10E-30		
2	3A8	195			collagen chain - mouse	6.50E-05
3	3A12	153			Keratin, ultra high sulfur matrix	2.80E-04
4	6E2	1295			HLA-B27 variant exon 3 (alpha2 domain)	5.10E-06
5	8A10	2200	Human erg protein (ets-related gene)	6.40E-34	transforming protein erg-2	1.20E-17
6	8E1	1940			NMDA receptor subunit NR2D	7.1E-04
7	8E4	1970			microtubule-associated protein 1A	1.90E-04
8	9E2	2188	EWS...erg (reciprocal translocation site)	1.10E-22	EWS-erg fusion protein type 1e	3.00E-13
9	9E8	2218	EWS...EWS-erg fusion protein type 9e	1.50E-36	transforming protein (ets-3) - fruit fly	7.30E-21
10	11A3	2260	human erg2 gene encoding erg2 protein	2.30E-35		
11	11A4	2254	human erg2 gene encoding erg2 protein	4.40E-25	ERG human transforming protein erg-e	8.30E-06
12	14F4	2025			phosphoglyceromutase (Drosophila)	2.40E-05
13	17D11	425	D. melanogaster single minded(sim) mRNA	3.30E-09	SIM drome single-minded protein	6.90E-10
14	SnsA0310	2182	EWS...erg (translocation, type 1e and 9e)	2.80E-63	EWS-erg fusion protein type 1e	1.10E-33
15	SnsD0320	2315			hypothetical protein (src 5' region)	9.30E-04
16	SnsB32	1253			interleukin-4 receptor alpha chain precursor	1.10E-03

DISCUSSION

The extensive amount of physical and genetic information and resources available for chromosome 21 was instrumental in allowing the construction of the next phase of higher resolution physical and transcript maps by us and others (11–13,18–20). Our approach was to: (i) use the YAC map to identify a subset of cosmids for a 2.5 Mb region; (ii) bin the cosmids according to regions of YAC overlap; (iii) assay the cosmids for the presence of known markers; (iv) isolate exons from the cosmid bins; and (v) use the isolated exons as markers to further order cosmids within and between bins. Integrating exon content mapping with genetic mapping information culminated in the relative placement of ordered cosmids covering ~88% (2.2 Mb/2.5 Mb) of this region with an average exon density of 1 every 25 kb, the identification of cosmid gaps and estimates of gap sizes. In this map, the average cosmid gap is estimated to be ~10 kb. Only one cosmid gap appears to be of significant size, ~100 kb, and current efforts are focused on closing it by cosmid end clone walking. A previously reported cosmid map of this region describes a minimal set of 76 cosmid clones and associated cDNA fragments (12,13). The addition of the information described here provides depth to the cosmid map, additional contig ordering and linking, and complements the transcript map with non-redundant gene fragments in the form of exons.

Localization of cDNA clones to this region by other groups has recently been reported (11,13). Comparison of these cDNA sequences with the exon sequences revealed one overlapping clone, a cDNA fragment corresponding to the ERG gene that contained exon 9E2. The YAC and cosmid location of the remaining cDNA clones places them in regions where other exons have been localized and therefore some of them may represent different fragments of the same genes. The combined efforts of exon amplification and cDNA selection have likely isolated fragments of many of the genes in this region. However, the coverage of the transcript map can only be as comprehensive as the initial cosmid map from which the exons and cDNAs are derived. Where there are cosmid gaps in the map, there are potential gene gaps and current efforts are focused on investigating these areas.

The blast analysis of the exons has identified 4.9% (5/102) with significant similarities [$p(N) > 10^{-4}$] to reported sequences, in addition to those exons from the ERG gene. This is comparable to the 5.5% (1/18) significant similarities reported for BLAST analysis of cDNA selected clones from this region (12) with analogous values for p(N). Possible explanations for this finding are that many of the genes in this region are novel or, more likely, that the amount of sequence available from the exons is insufficient to identify similarities with known genes in many instances. As additional sequence from the same gene is determined from cDNA analysis more relationships with known genes might be expected to emerge.

Another application of the high resolution map of this region will be in the area of fine mapping of sequence copy number in some individuals with partial aneuploidies for chromosome 21. By using individual cosmid clones as FISH probes, the requirement for informative polymorphisms to determine sequence copy number is alleviated and the analysis can be performed approximately every 50 kb between CBR and ERG with the existing cosmids. This will allow refinement of the definitions of some phenotype-genotype correlations in DS. Having cloned exons available from these cosmids provides immediate access to the genes that contribute to the DS phenotypes mapped to that region.

It is expressly because the definition of DS regions is complicated that large scale gene isolation efforts are needed since many genes from several regions of the chromosome will likely need to be studied in order to decipher their roles in DS. In addition, the experience gained from utilizing large scale gene isolation approaches in the DS region should allow them to be more effectively applied to other genomic regions of interest. The challenge of the future is to correlate over-expression of these genes, singly or in combination, with the presence of a DS phenotype.

MATERIALS AND METHODS

YAC clones

DNA from YAC clones representing the region between CBR and ERG (21,22) was made as described (23). Sizes of the YACs were determined by hybridization of Southern blots of pulsed field gels to human *Cor1* DNA. The presence of STS in the YACs were confirmed by PCR using primer sequences available from public databases.

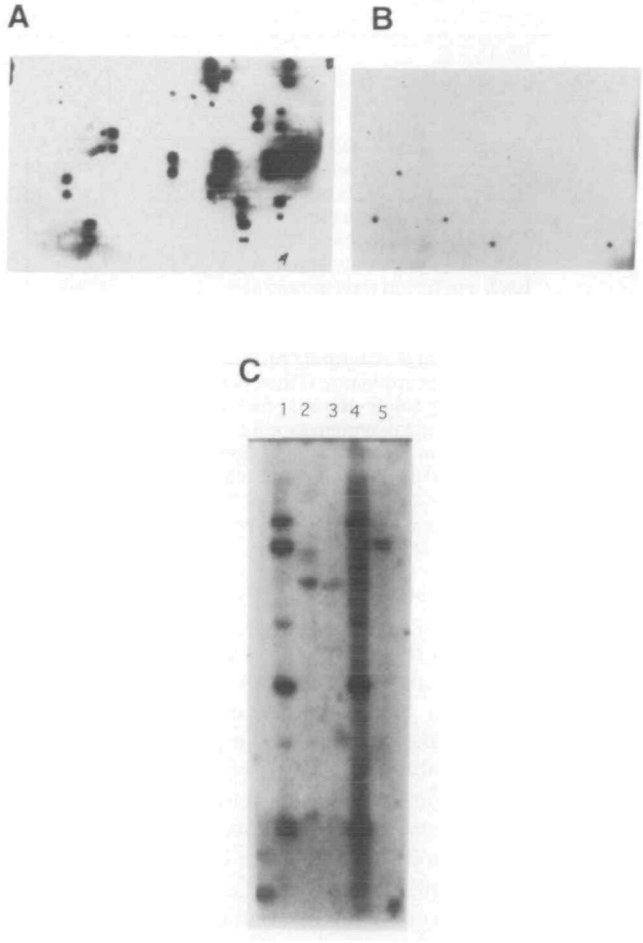

Figure 4. Localization of the cDNA clone containing exon 3E2 to chromosome 21. (**A**) Cosmid colony grid. Colony grids of the 394 cosmids identified by hybridization with YAC *Alu*-PCR products were prepared using a Biomek 1000 robot. Clones were stamped in duplicate in a 4×2 array. Cosmids containing the cDNA are indicated by the duplicate signals. This cDNA is present in 17 cosmids, five of which also contain exon 3E2 (Table 1). (**B**) Colony grid of exon clones. Exons isolated from cosmid pools were cloned, arrayed in 96-well dishes and stamped on nylon filters. This cDNA hybridizes to exon 3E2, which was used to isolate the cDNA, and four other trapped exons. (**C**) Genomic Southern blot. Digested genomic DNA, isolated from cell lines, was separated on 1.0% agarose gels, transferred to nylon membranes. Lanes are the same as described in Figure 2B. The bands common to lanes 1 and 4 indicate this cDNA localizes to chromosome 21. Similarly, a common band in lanes 2 and 5 indicate cross hybridization to mouse genomic DNA and localization to mouse chromosome 16. This sequence is also conserved in hamster genomic DNA as indicated by the band in lanes 2 and 3.

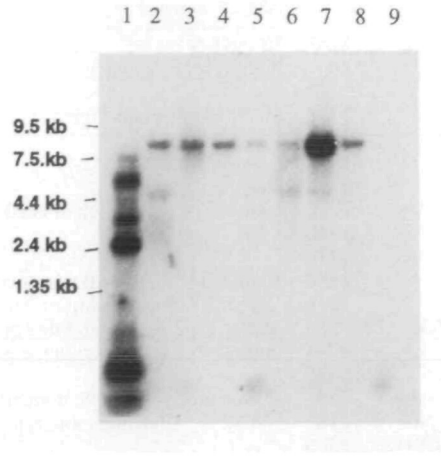

Figure 5. Northern blot analysis with the cDNA clone containing exon 3E2. Tne northern blot containing poly-A^+ mRNA from human adult tissues was obtained from Clontech and hybridized with the gel purified cDNA insert from the clone containing exon 3E2. Lanes: 1, marker; 2, heart; 3, brain; 4, placenta; 5, lung; 6, liver; 7, sk. muscle; 8, kidney; 9, pancreas. This cDNA hybridizes to an mRNA of ~8.0 kb that is expressed in all tissues present.

Isolation of cosmid clones

Cosmid clones were identified by hybridization of colony grids with gel purified YAC DNA or YAC *Alu*-PCR products (Fig. 1). When *Alu*-PCR products were used, DNA from YAC clones was amplified with four different *Alu* primers, combined and hybridized to colony grids (3×3 or 4×4) of a flow sorted chromosome 21 cosmid library, LL21NC02, constructed at Lawrence Livermore National Laboratory. Positive cosmids were picked into microtitre plates and colony grids were prepared. *Alu*-PCR products from individual YACs were individually hybridized to this subset of cosmid clones. Cosmid clones that rescreened positive with individual YAC *Alu*-PCR products were placed in bins. Cosmid bins are defined by regions of overlap and non-overlap between different YACs. The individual cosmids in each bin are listed in Table 1. Miniprep DNA was prepared and the restriction patterns of cosmids within a bin were compared to identify identical cosmids. Non-identical cosmids were chosen for subsequent exon amplification experiments.

Isolation of exons

Subsets of non-identical cosmids from each bin were subjected to exon amplification (13; Gibco-BRL). A total of 218 cosmids were trapped in 21 pools of up to 15 clones per pool. Forty eight exon clones from each cosmid pool (1008 total exon clones) were picked to 11 microtitre plates. Initially, 12 exon clones, (252 total clones) were sequenced from each pool to identify unique exons and background products, which consist of repetitive elements and vector sequences. Redundant clones in each trapping pool were identified by hybridizing PCR amplified products from these 12 exons and background containing clones to colony grids of all the clones obtained from the trapping experiment. A second round of exon trapping was performed using 56 cosmids that either yielded no exons in the initial trapping experiments, or were obtained as the result of chromosome walking or PCR screening of the cosmid library. The second round of trapping was performed on individual clones or on pools of up to seven cosmids per pool. The resulting exons were analyzed as described above for the initial trapping experiments. A total of 404 potentially unique exon clones were identified. Full sequence has been obtained for 115 unique exon clones using the method of Sanger (24) and a sequencing kit (USB). Sequence database analysis was performed using the BLAST network service of the National Center for Biotechnology Information. The Genbank accession numbers for the exons included in Figure 3 are X85338-X85367 for SNS exons and R82098-R82175 for exons beginning with 'e'.

PCR conditions

The primers used in the *Alu*-PCR of YAC DNA, derived from the *Alu* sequence, were: *Alu* S (5′-gaggttgcagtgagccgagat-3′), *Alu* J (5′-gctgcagtgagccgtgat-3′), ALE 1 (5′-gcctcccaaagtgctgggattacag-3′), ALE 3 (5′-ccactgcactccagcctggg-3′), END-2 (5′-acagagc/tg/aagactccgtctc-3′), END-1 (5′-acagagcgagactcc/tg/atc-3′). The *Alu* S and *Alu* J primers are combined in a single PCR reaction as are the END-1 and END-2 primers. The Ale 1 and Ale 3 primers are used individually. Amplification conditions were 94°C, 30 s; 60°C, 30 s; 72°C, 2 min for 35 cycles. Amplification of YAC DNA with STS was performed using primer sequences described in public databases and conditions as above with a 1 min extension time. Cloned exon products were amplified using vector sequences SD2 and SD2R. Purified amplification products were used as hybridization probes in subsequent mapping experiments.

Hybridization conditions

Probes for hybridization to colony grids or genomic Southern blots were exons ~50 bp or longer, cDNA clones or *Alu*-PCR products. Probes were labelled by random priming with ^{32}P-dATP and ^{32}P-dCTP and pre-annealed

with human placental and/or *Cot*1 DNA (BRL) if necessary. Preannealed probes were mixed with 50 μg *Cot*1 DNA or 5 μg *Cot*1 and 2 mg of placental DNA in 5× SSC, 0.1% SDS, heated to 100°C for 5 min followed by incubation at 65°C for 1 or 4 h, respectively, before adding to hybridization buffer. Hybridizations were performed in Church and Gilbert hybridization buffer at 65°C overnight. Filters were washed in 2× SSC, 0.1% SDS at room temperature followed by 1× SSC, 0.1% SDS at 65°C and 0.5× SSC, 0.1% SDS at 65°C for 15–20 min each, except for *Alu*-PCR product probes, which were washed to 0.1× SSC, 0.1% SDS at 65°C for 15–20 min. Northern blots were hybridized in 50% formamide, 10% dextran sulfate, 1 M NaCl, 1× Denhardt's solution, 50 mM Tris (pH 7.5), 0.5% SDS, 100 μg/ml sheared denatured salmon sperm DNA at 42°C and washed to a stringency of 0.5× SSC, 0.1% SDS. Filters were exposed to autoradiographic film from several hours (colony grid hybridizations) to several days (genomic Southern and northern blots).

ACKNOWLEDGEMENTS

The authors would like to thank C.J.Stotler and J.Rutter for excellent technical support and M.Lalioti and J-L.Blouin for their assistance with certain experiments. This work was supported in part by NIH grant NS32675 to MKM and by the Swiss FNRS grant 31–33965.92, 31–40500.94 and EC grant GENE-CT93–0015 to SEA.

REFERENCES

1. Lejeune, J., Gautier, M. and Turpin, R. (1959a) Etude des chromosomes somatiques de neuf enfants mongoliens (Study of the somatic chromosomes of nine mongol children) *C. R. Acad. Sci. (Paris)* **248**: 1721–1722.
2. Lejeune, J., Turpin, R. and Gautier, M. (1959b) Le mongolisme, maladie chromosomique (Trisomie). (Monglosim, a chromosomal disease (Trisomy)). *Bull. Acad. Nat. Med.* **143**: 256–265.
3. Summit, R.L. (1981) *Trisomy 21 (Down syndrome) Research Perspectives.* (1981) (FF de la Cruz and PS Gerald, eds). pp 225–235. University Park Press, Baltimore.
4. McCormick, MK., Schinzel, A., Petersen, M.B., Stetten, G., Driscoll, D.J., Cantu, E.S., Tranebjaerg, L., Mikkelsen, M., Watkins, P.C. and Antonarakis, S.E. (1989) Molecular genetic approach to the characterization of the 'Down syndrome region' of chromosome 21. *Genomics* **5**: 325–331.
5. Delabar, J-M., Theophile, D., Rahmani, Z., Chettouh, Z., Blouin, J-M., Prieur, M., Noel, B. and Sinet, P-M. (1993) Molecular mapping of twenty-four features of Down syndrome on chromosome 21. *Eur. J. Hum. Genet.* **1**: 114–124.
6. Sinet, P-M., Theophile, D., Rahmani, Z., Chettouh, Z., Blouin, J-L., Prieur, M., Noel, B. and Delabar, J-M. (1993) Molecular mapping of the Down syndrome phenotype on chromosome 21 in *The phenotypic mapping of Down syndrome and other aneuploid conditions.* Wiley-Liss, Inc. p. 63–86.
7. Rahamni, Z., Blouin, J.L., Creau-Goldberg, N., Watkins, P.C., Mattei, J.F., Poisonnier, M., Prieur, M., Chettouh, Z., Nicole, A., Aurias, A., Sinet, P-M. and Delabar, J-M. (1989) Critical role of the D21S55 region on chromosome 21 in the pathogenesis of Down syndrome. *Proc. Natl. Acad. Sci. USA* **86**: 5958–5962.
8. Korenberg, J.R., Kojis, T.L., Bradley, C. and Disteche, C. (1989) Down syndrome and band 21q22.2: molecular definition of the phenotype. *Am. J. Hum. Genet.* **45**: (Suppl) A79.
9. Korenberg, J.R., Kawashima, H., Pulst, S-M., Ikeuchi, T., Ogasawara, N., Yamamoto, K., Schonberg, S.A., *et al.* (1990) Molecular definition of a region of chromosome 21 that causes features of the Down syndrome phenotype. *Am. J. Hum. Genet.* **47**: 236–246.
10. Korenberg, J. R., Chen, X-N., Schipper, R., Sun, Z., Gonsky, R., Gerwehr, S., Carpenter, N., Daumer, C., Dignan, P., Disteche, C., Graham, Jr., J. M., Hugdins. L., McGillivray, B., Miyazaki, K., Ogasawara, N., Park, J. P., Pagon, R., Pueschel, S., Sack, G., Say, B., Schuffenhauer, S., Soukup, S. and Yamanaka, T. (1994) Down syndrome phenotypes: The consequences of chromosomal imbalance. *Proc. Natl. Acad. Sci. USA* **91**: 4997–5001.
11. Cheng, J-F., Boyartchuk, V. and Zhu, Y. (1994) Isolation and mapping of human chromosome 21 cDNA: Progress in constructing a chromosome 21 expression map. *Genomics* **23**: 75–84.
12. Patil, N., Peterson, A., Rothman, A., de Jong, P. J., Myers, R. M. and Cox, D. R. (1994) A high resolution map of 2.5 Mbp of the Down syndrome region on chromosome 21. *Hum. Mol. Genet.* **3**: 1811–1817.
13. Peterson, A., Patil, N., Robbins, C., Wang, L., Cox, D. R. and Myers, R. M. (1994) A transcript map of the Down syndrome critical region on chromosome 21. *Hum. Mol. Genet.* **3**: 1735–1742.
14. Buckler, A.J., Chang, D.D., Graw, S.L., Brook, J.D., Haber, D.A., Sharp, P.A. and Housman, D.E. (1991) Exon Amplification: A strategy to isolate mammalian genes based on RNA splicing. *Proc. Natl. Acad. Sci. USA* **88**: 4005–4009.
15. Crete, N., Gosset, Ph., Theophile, D., Duterque-Coquillaud, M., Blouin, J. L., Vayssettes, C., Sinet, P-M. and Creau-Goldberg, N. (1993) Mapping the Down syndrome chromosome region. *Eur. J. Hum. Genet.* **1**: 51–63.
16. Nambu, J. R., Lewis, J. O., Wharton, K. A. and Crews, S. T. (1991) The Drosophila single-minded gene encodes a helix-loop-helix protein that acts as a master regulator of CNS midline development. *Cell* **67**: 1157–1167.
17. Muralidhar, M. G., Callahan, C. A. and Thomas, J. B. (1993) Single-minded regulation of genes in the embryonic midline of the Drosophila central nervous system. *Mech. Dev.* **41**: 129–138.
18. Antonarakis, S.E. (1993) Human chromosome 21: genome mapping and exploration. *Trends Genet* **9**: 142–148.
19. Nizetic, D., Gellen, L., Hamvas R. M. J., Mott, R., Grigoriev, A., Vatcheva, R., Zehetnet, G., Yaspo, M-L., Dutriaux, A., Lopes, C., Delabar, J-M., Van Broeckhoven, C., Potier, M-C. and Lehrach, H. (1994) An integrated YAC-overlap and 'cosmid-pocket' map of the human chromosome 21. *Hum. Mol. Genet.* **3**: 759–770.
20. Soeda, S., Hou, D-X., Osoegawa, K., Atsuchi, Y., Yamagata, T., Shimokawa, T., Kishida, H., Soeda, E., Okano, S., Chumakov, I., Cohen, D., Raff, M., Gardiner, K., Graw, S., Patterson, D., De Jong, P., Ashworth, L., Slezak, T. and Carrano, A. V. (1995) Cosmid assembly and anchoring to human chromosome 21. *Genomics* **25**: 73–84.
21. Chumakov, I. M., Le Gall, I., Billault, A., Ougen, P., Soularue, P., Guillou, S., Rigault, P., Bui, H., De Tand, M-F., Barillot, E., Abderrahim, H., Cherif, D., Berger, R., Le Paslier, D. and Cohen, D. (1992) Isolation of chromosome 21-specific yeast artificial chromosomes from a total human genome library. *Nature Genet.* **1**:222–225.
22. Cohen, D., Chumakov, I., Weissenbach, J. (1993) A first-generation physical map of the human genome. *Nature* **366**:698–701.
23. Davis, R. W., Thomas, M., Cameron, J., St. John, T P., Scherer, S. and Padgett, R. A. (1980) Rapid DNA isolation for enzymatic and hybridiaztion analysis. *Meth. Enzymol.* **65**:404–411.
24. Sanger, F., Nicklen, S. and Coulsen. (1977) *Proc. Natl. Acad. Sci. USA* **74**: 5463–5467.
25. Dufresne-Zacharia, M-C., Dahmane, N., Theophile, D., Orti, R., Chettouh, Z., Sinet, P-M. and Delabar, J-M. (1994) 3.6 Mb genomic and YAC physical map of the Down syndrome chromosome region on chromosome 21. *Genomics* **19**: 462–469.
26. McCormick, M. K., Buckler, A., Bruno, W., Campbell, E., Shera, K., Torney, D., Deaven, L. and Moyzis, R. (1993) Construction and characterization of a YAC library with a low frequency of chimeric clones from flow sorted human chromosome 9. *Genomics* **18**: 553–558.