Am. J. Hum. Genet. 51:1251–1264, 1992

Analysis of Chromosome 21 Yeast Artificial Chromosome (YAC) Clones

Flora Tassone,*,† Suzanne Cheng,* and Katheleen Gardiner*

*Eleanor Roosevelt Institute for Cancer Research, Denver; and †Institute of Human Genetics, "A. Gemelli" School of Medicine, Rome

Summary

Chromosome 21 contains genes relevant to several important diseases. Yeast artificial chromosome (YAC) clones, because they span >100 kbp, will provide attractive material for initiating searches for such genes. Twenty-two YAC clones, each of which maps to a region of potential relevance either to aspects of the Down syndrome phenotype or to one of the other chromosome 21–associated genetic diseases, have been analyzed in detail. Clones total ~6,000 kb and derive from all parts of the long arm. Rare restriction-site maps have been constructed for each clone and have been used to determine regional variations in clonability, methylation frequency, CpG island density, and CpG island frequency versus gene density. This information will be useful for the isolation and mapping of new genes to chromosome 21 and for walking in YAC libraries.

Introduction

Chromosome 21 is the smallest of the human chromosomes and is postulated to be relatively low in gene content. Nevertheless, genes contained within the long arm of the chromosome appear to be involved in a number of medically, biologically, and socially significant diseases. These include familial Alzheimer disease (FAD), amyotrophic lateral sclerosis (ALS), progressive myoclonus epilepsy (PME), holoprosencephaly (HP), and Down syndrome (DS) (Pueschel 1982; Epstein 1986; St George-Hyslop et al. 1987; Estabrooks et al. 1990; Lehesjoki et al. 1991; Siddique et al. 1991).

Candidate genes are not available for most of these diseases (but see Goate et al. 1991; Schellenberg et al. 1991); however, a search for such genes is feasible for two reasons. First, the genetic and physical maps of the chromosome are quite advanced (Cox et al. 1990; Gardiner et al. 1990*b;* Owen et al. 1990; Burmeister et al. 1991), and the DS phenotypic map continues to be refined (McCormick et al. 1989; Rahmani et al. 1989; Korenberg et al. 1990, 1992; Korenberg 1991). As a result, a region of interest can generally be limited to a few megabases or less, containing several known probes. Second, yeast artificial chromosome (YAC) clones are becoming available and widely distributed for much of the chromosome (Patterson 1991). Hence, it is relatively straightforward to obtain a YAC or YACs, potentially containing relevant genes, that map to a region of interest.

We have selected 22 YAC clones from human chromosome 21, obtained from the St. Louis human library under the auspices of the International Joint YAC Screening Effort (Patterson 1991). Together, these clones comprise >6 megabases (Mbp) of DNA (15% of the long arm) and derive from all regions of the chromosome. Each YAC is well placed with respect to candidacy for a specific disease or aspect of DS. We have constructed pulsed-field restriction maps for these clones and used them to investigate regional variations in (1) distribution of unique sequences currently on the map, (2) patterns in methylation frequency, (3) rare restriction-site and potential CpG island distribution, and (4) gene size and presence of CpG island. This analysis should aid in choosing gene isolation procedures and in pursuing investigations into human genome organization.

Material and Methods

YAC Clones

YAC clones listed in tables 1 and 2 (except for those for D21S65) were identified by PCR screening of the St. Louis human YAC library (Brownstein et al. 1989;

Received March 25, 1992; final revision received August 19, 1992.

Address for correspondence and reprints: Katheleen Gardiner, Ph.D., Eleanor Roosevelt Institute, 1899 Gaylord Street, Denver, CO 80206.

0002-9297/92/5106-0008$02.00

Table 1

YAC Clones with Limited Analysis

YAC	Probe	Size (kb)	Reason for Not Pursuing
B231C4	D21S120	120	Larger alternative available
B126C4	D21S120	200	Larger alternative available
A120D9	D21S120	260	Larger alternative available
B148E8	D21S120	280	Larger alternative available
B254A9	D21S120	410	Possibly identical
B207B4	D21S12	100	Larger alternative available
A233D1	D21S12	100	Larger alternative available
A47C3	GART	350	Chimeric
A62H12	D21S65	320	Larger alternative available
A200E11	D21S3	900	Chimeric
B136H7	D21S13	385	Map divergence[a]
A168D10	APP	335	Map divergence[a]
D56F5	D21S111	620	Map divergence;[a] chimeric (FISH)
D112A5	SOD1	185	Small and chimeric (R,v)[b]
A165D7	D21S65	200	Within C14B2
C4C10	D21S65	250	Within C14B2
B119G7	D21S65	1,300	Chimeric; 250 kb within C14B2
B234B10	D21S55	185	Map divergence[a]

NOTE.—Listed are the YAC clones discarded after preliminary analysis.

[a] Pulsed-field map significantly different from that obtained for other, nonchimeric YACs containing the same probe (for other YACs, see table 2 and fig. 3).

[b] v = Vectorette PCR; and R = right end.

Green and Olson 1990). YACs for SOD1 and GART have been reported (Imai and Olson 1990; Gnirke et al. 1991). YACs for D21S120, D21S111, and APP (D110G6) were obtained by B. Brownstein (Washington University, St. Louis). The remainder were obtained by the Chromosome 21 International Joint YAC Screening Effort (headed by D. Patterson, Eleanor Roosevelt Institute). YACs containing D21S65 were identified by F. Tassone using the pulsed-field gel/Southern blot method (Mendez et al. 1991). All clones were routinely grown in selective medium to assure maintenance of the YAC. YAC DNA was prepared, as previously described, in agarose plugs at a concentration of 7×10^8 cells/ml. To determine whether the ends of the YAC clones mapped to chromosome 21, three different procedures were used. Vector-Alu PCR with primers TC65, 517, and 278 (Nelson et al. 1989); vectorette PCR analysis (Riley et al. 1990) using *Bgl*II, *Bam*HI, and *Bcl*I; left-end rescue methods (Burke et al. 1987); and fluorescent in situ hybridization (FISH) (Trask et al. 1989) were all used essentially as described. End clones for the ETS2 YAC were obtained and mapped to chromosome 21 by M. Brennan (National Institute of Mental Health); FISH with the CD18 YAC was done by W.-L. Kuo and J. Gray (University of California at San Francisco) and with the D21S17 YAC by J. Gingrich (Lawrence Berkeley Laboratory). Both size of the YAC containing D21S120 and the presence of D21S48 and D21S16 were originally determined by S. Rider (UCSF). YAC addresses are given in tables 1 and 2. For brevity, YACs analyzed in detail (table 2) are referred to by the name of the probe used to identify the clone, e.g., yD21S65 is clone C14B2 and contains probe D21S65. Where two YACs with the same probe are discussed, they are designated as "L" and "S," for the larger and smaller, respectively.

Pulsed-Field Gel Electrophoresis (PFGE)

Sizing of YAC chromosomes and separation of complete digests were done on the TAFE system (Gardiner et al. 1990*b*); partial digests were separated in the ED system, a modification of the original Schwartz device (Schwartz et al. 1989). Both gel types used 0.25 × TBE buffer (10 × TBE = 0.9 M Tris-borate/20 mM EDTA). Electrophoresis conditions are given in the figure legends. Size markers were concatamers of lambda (48.6 kb), the chromosomes of the host *Saccharomyces cerevisiae* strain AB1380, and the 90–

Table 2

Analysis of YAC Clones

YAC[a]	Probe[b]	Other[c]	Size (kb)	Region[d]	Chimerism[e]
B251H6	D21S120	D21S16, D21S48, and Not50	410	I	Yes (f)
B236A6	D21S12		250	I	No (f; R,v)
A109F12	D21S13		265	I	No (f)
A210B6	D21S52	D21S59	520	I	Yes (v)
B47F7	D21S52		150	I	?
A300D8	D21S110		310	I	No (f)
B2B3	D21S11	D21S1	500	I	Yes (f)
D110G6	APP	Eag118-1	835	I	No (v)
A151E3	D21S12	D21S99 and D21S111	350	I	No (v)
A228D8	D21S93		265	I	No (f; L,v)
D142H8	GART	SON	600	II	No (ep)[f]
C14B2	D21S65		450	II	No (ep and va)
A222A12	D21S17		270	III	No (f)[f]
B134B9	D21S55		330	III	No (v)
A125B12	ERG	D21S60	285	III	No (v)
B19C12	ERG		145	III	No (v)
A196B6	ETS2		250	IV	No (ep)[f]
B258F4	D21S15		165	IV	No?
B208G3	D21S64	D21S53	300	IV	No[g]
B45F3	D21S19		210	IV	No (f)
B45A8	CBS	CRYA	125	IV	No?[g]
A228B4	CD18		330	IV	No (f)

[a] Address in the St. Louis library.
[b] D locus used to isolate the YAC.
[c] Additional probes found in the YAC.
[d] See fig. 2.
[e] f = FISH; v = vectorette PCR; ep = cloning of end probes; va = vector Alu PCR; and R = right end and L = left ends, respectively.
[f] Information from personal communications by A. Gnirke (GART) (Gnirke et al. 1991), J. Gingrich (D21S17), M. Brennan (ETS2), and W.-L. Kuo and J. Gray (CD18).
[g] Not likely to be chimeric, because of the positions of unique sequences within the YAC. For discussion of chimerisms, see the text.

1,600 kb chromosomes of *S. cerevisiae* strain yPH149 (provided by P. Hieter).

Enzyme Digestions

*Not*I, *Sfi*I, and *Nru*I were purchased from Stratagene; *Sst*II from BRL; and others from New England Biolabs. Complete digests were done according to a method described elsewhere (Gardiner et al. 1990*b*) by using ~30 units of enzyme. Partial digests were done according to a method described elsewhere (Burke et al. 1987), for 3 h and using three or four different concentrations of enzyme, generally ranging from 0.01 to 20 units, depending on activity and lot number of the enzyme. The map for yGART is from Gnirke et al. (1991), with the addition of *Eag*I and *Sst*II sites.

Probes

Chromosome 21 unique sequence probes are listed in table 1 and are from the collection used in previous studies (Gardiner et al. 1990*b*). RT1 is a full-length cDNA clone, containing the entire coding region and the 3′ untranslated segment, for the 751 form of the APP gene (a gift from R. Tanzi); RT1-1.1 is a 1.1-kb *Eco*RI fragment of this clone, containing the 3′ sequences found in the original APP cDNA, FB63 (Tanzi et al. 1987); RT1-2.1 contains the 5′ sequences. CH49/39 is a 2.7-kb *Pst*I fragment from cosmid 49 (Brahe et al. 1990). Sequence analysis (F. Tassone, unpublished data) indicates that it contains the promoter of the APP gene (Salbaum et al. 1988). HO8 is a 3.5-kb *Hin*dIII fragment from cosmid ICRFc-102H08124, isolated from the ICRF library (Lehrach et al. 1990) by screening with D21S65. SON3 is a cDNA mapping to 21q22.1 (Berdichevskii et al. 1988). LambdaNot50Pvu3 is a 4-kb *Eco*RI/*Pvu*I fragment, and lambdaEag118-1 is a 1-kb *Eco*RI fragment; both are from lambda clones previously reported (Gao et al. 1991*b*). The pYAC4 left- and right-arm vector–

specific pBR fragments used in analysis of partial digestions have been described elsewhere (Burke et al. 1987). The 2.6-kb fragment is specific for the trp (left) arm of the vector, the 1.6 for the ura (right) arm. Probes were labeled by the random primer method to specific activities of 10^8–10^9 cpm/μg (Feinberg and Vogelstein 1989).

Results

Preliminary YAC Analysis

Forty YAC clones from the St. Louis human library, representing screening for 22 probes, were subjected to all or part of the following analysis: (1) determination of size, by pulsed-field analysis; (2) examination of probe content, based on current physical maps; (3) investigation of chimerism, by end-cloning, vectorette PCR, vector-Alu PCR, or FISH; and (4) construction of pulsed-field maps, using nine rare-cutting restriction enzymes. Only the largest YAC obtained with each probe was pursued initially; further YACs were added as chimerism or other ambiguities made the first choice less attractive. On this basis, the YACs have been divided into two groups. Table 1 lists those for which only limited analysis was done, and it gives the reason for not pursuing each further. Table 2 gives detailed information for the 22 YACs on which detailed analysis has focused. These YACs represent screening for 19 probes. All clones appeared stable during prolonged growth in selective medium.

Of the clones listed in table 2, 17 are nonchimeric; the smallest is 125 kb, and the largest is 835 kb. For several YACs, lack of chimerism was determined by FISH, and it is true that this method may not detect small segments mapping elsewhere. If some of these YACs are indeed chimeric, it is likely to be a small region and does not affect the further analysis. YACs were sequentially hybridized with total human DNA, with the probe used to screen the library, and with additional probes previously mapped to the same region. Table 2 gives the information on size, probe content, and chimerism.

Nine YACs contained one to three additional probes. For most of these, proximity had already been established by pulsed-field linkage (Gardiner et al. 1990*b*); however, this was often on fragments of considerable size (e.g., D21S52 and D21S59 on fragments approaching 2 Mbp, ERG and D21S60 on >800 kb, etc.). Thus, YAC analysis has refined the physical mapping data for many of these probes. In addition, it also points out that this collection of probes is perhaps even more closely clustered than previously expected, in particular, in 21q21 (see fig. 2 and Discussion).

Figure 1 shows the location of each YAC along the chromosome. Positioning is based on physical mapping data previously obtained for probes within the YACs (Gardiner 1990; Gardiner et al. 1990*b*). Regions I–IV had also been determined, using pulsed-field and base-compositional data, to have the following characteristics: region I (21cen-proximal 21q22.1)—AT rich and gene poor; region II (distal 21q22.1)—CG rich and gene rich; region III (21q22.2)—AT rich and gene rich; and region IV (21q22.3), very CG rich and very gene rich (Gardiner 1990; Gardiner et al. 1990*a*, 1990*b*). Regional localization of the YACs is of interest in considering possible disease gene associations and for predicting possible CpG island and gene densities.

Construction of Pulsed-Field Maps

To identify clusters of rare sites that may indicate CpG islands (and thus the 5′ end of gene sequences) and to obtain information relevant to methylation patterns, pulsed-field maps were constructed for each of the 22 clones. Nine enzymes that cut infrequently in human DNA were used: *Not*I, *Bss*HII, *Mlu*I, *Nru*I, *Eag*I, *Sst*II, *Cla*I, *Sal*I, and *Sfi*I. These enzymes were chosen because they have been used in genomic pulsed-field analysis of the same probes (Gardiner et al. 1990*b*).

Each YAC was digested to completion with all nine enzymes and hybridized sequentially with total human, left- and right-end vector-specific sequences, and appropriate unique sequence probes. Partial digestions were carried out with each YAC for each enzyme with more than one site, and hybridizations again were done sequentially with left-end, right-end, and unique sequence probes. For YACs up to ~400 kb, sites were confirmed from both ends of the YAC. For small YACs (250 kb), this could be done with a single pulse time; for larger YACs, short pulse times were used for sites near the ends, and longer pulse times were used for the more distant sites. For YACs >450 kb (GART, D21S52, D21S11, and APP), the verification of sites by overlap could be done only for the central region. Examples are shown in figure 2. Notable features include the small number of *Sal*I and *Cla*I sites in the relatively large (410 kb) yD21S120 YAC (fig. 2*a*), the suggestion of a cluster of *Bss*HII and *Eag*I sites in yD21S65 (fig. 2*b*), and the large number of *Mlu*I sites in the 120-kb yCBS (fig. 2*c*) and *Not*I sites in yCD18 (fig. 2*d*).

Maps for all 22 YACs, with the nine enzymes, are shown in figure 3, grouped by chromosomal region and listed in order from centromere to telomere. In examining these maps, there are several points to note. (1) In region I, yD21S120 appears to be chimeric by

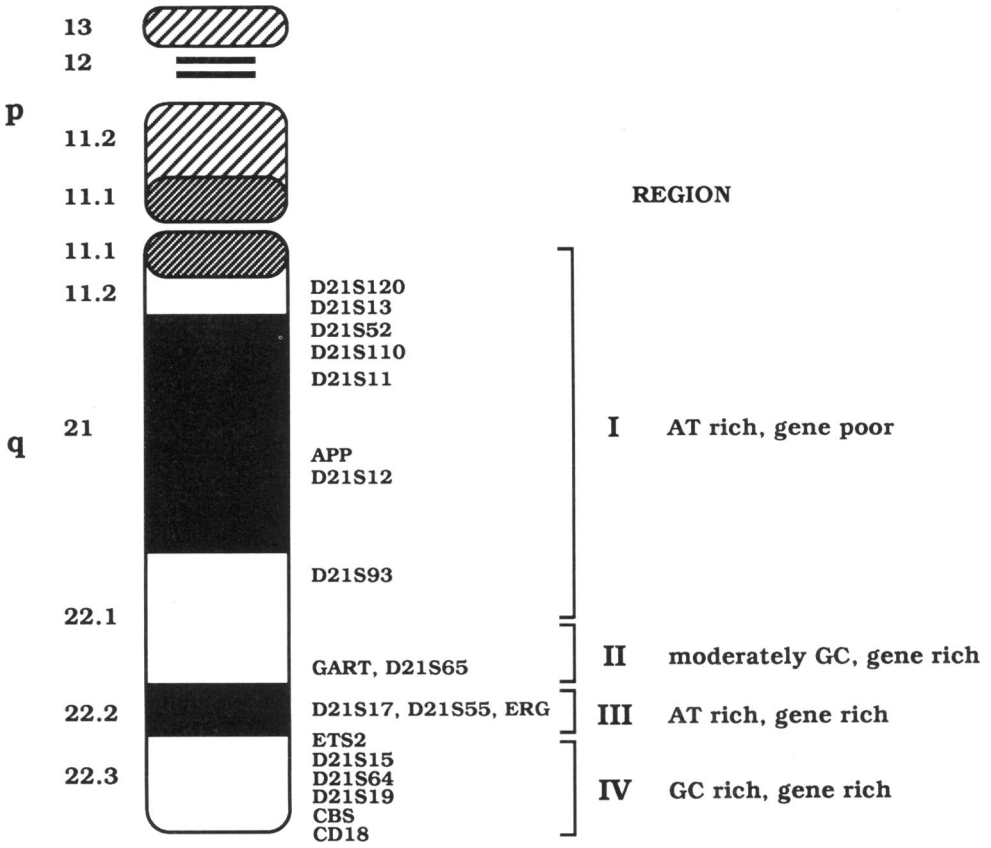


Figure 1 Regional localization of YAC clones. YAC clones (from table 2) are listed to the right of the chromosome schematic. Positions are based on the physical mapping of probes that each contains (Gardiner 1990; Gardiner et al. 1990*b*). Regions I–IV are defined from data reported by Gardiner et al. (1990*a*, 1990*b*), as described in the text.

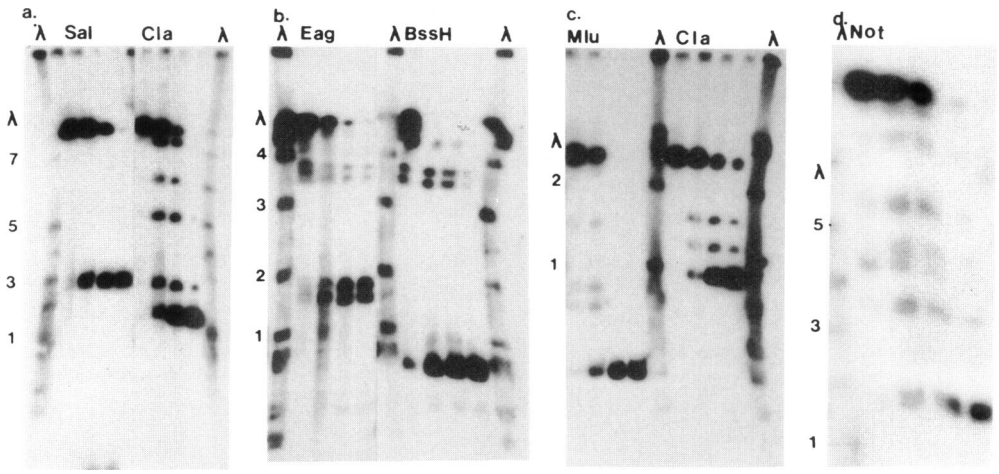
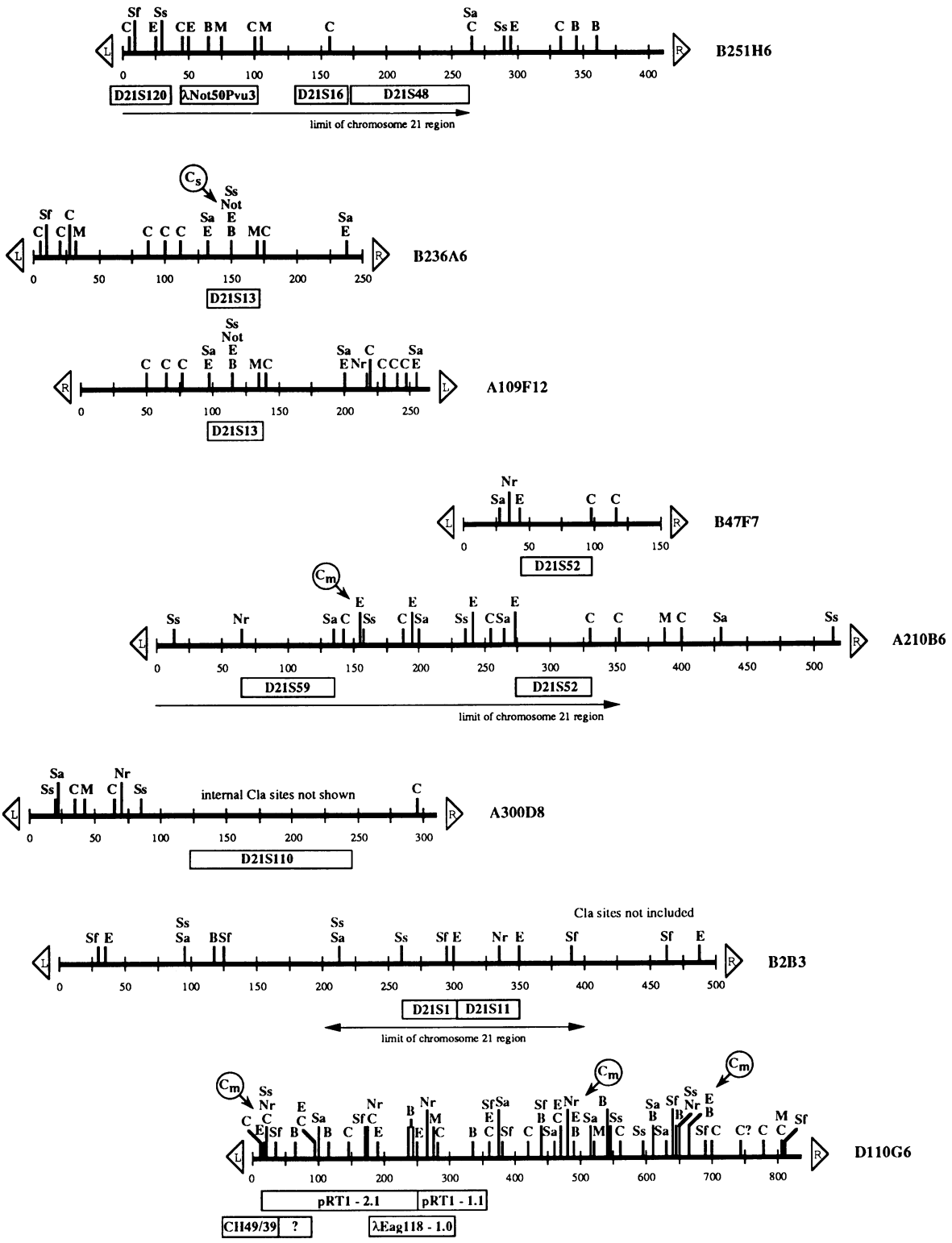
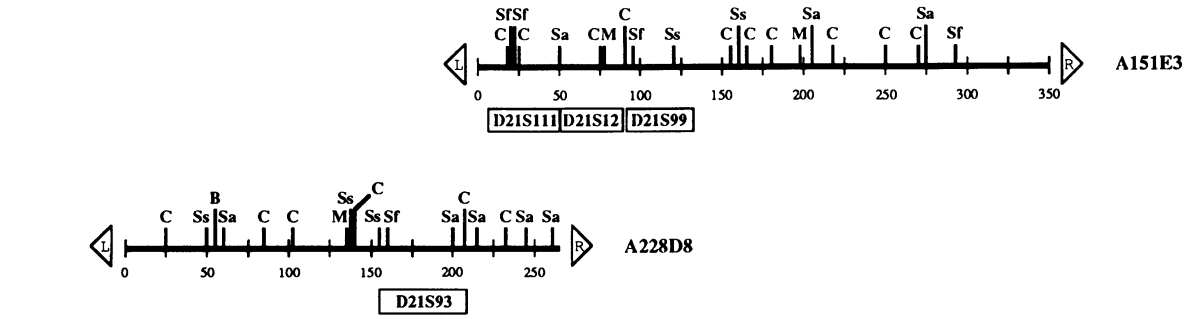
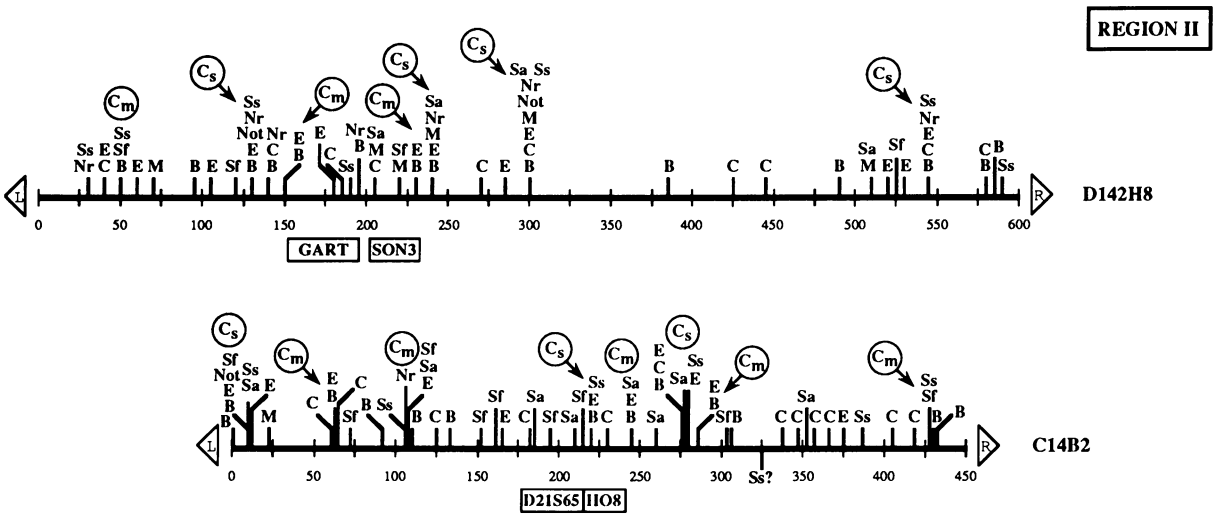


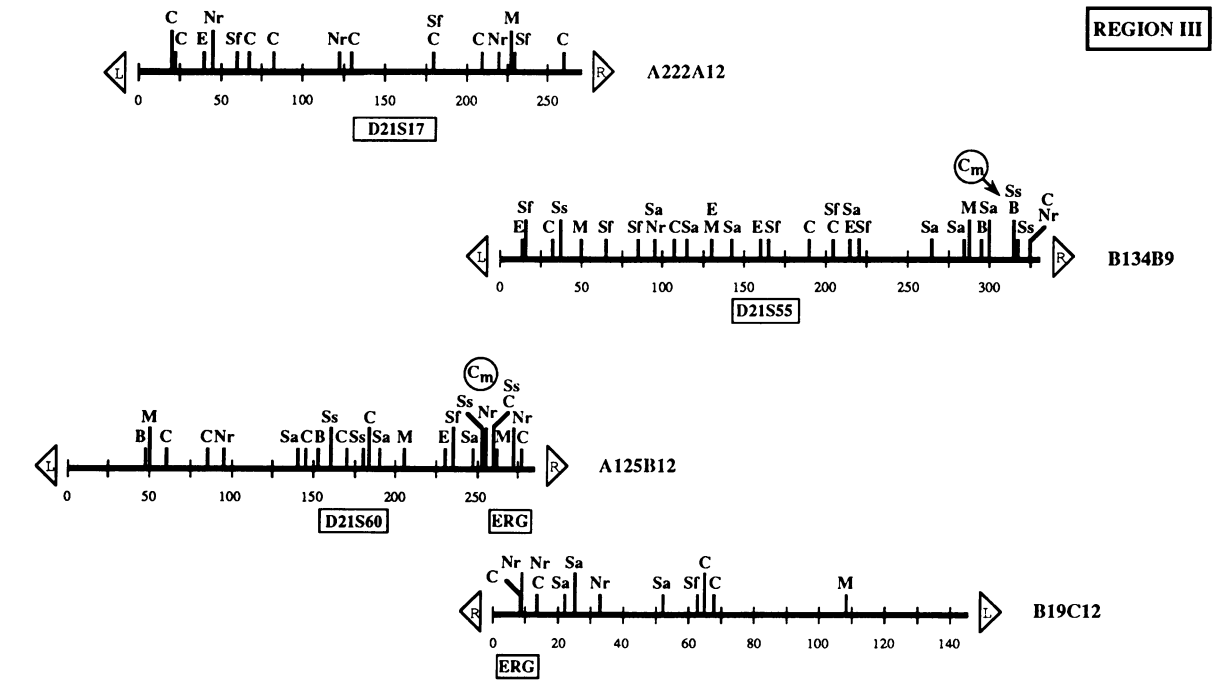
Figure 2 Partial digest analysis of YAC clones. YAC clones were digested with increasing concentrations of the enzymes indicated. Electrophoresis was at 200 V in the ED pulsed-field system, using 1% agarose. Hybridizations were with the vector-specific right-end probe. Size markers are given in concatamers of lambda. *a*, yD21S120, 20 s for 30 h. *b*, yD21S65, 7 s for 24 h. *c*, yCBS, 4 s for 24 h. *d*, yCD18, 15 s for 28 h.

REGION I


B251H6
D21S120
λNot50Pvu3
D21S16
D21S48
limit of chromosome 21 region
B236A6
D21S13
A109F12
D21S13
B47F7
D21S52
A210B6
D21S59
D21S52
limit of chromosome 21 region
A300D8
internal Cla sites not shown
D21S110
B2B3
Cla sites not included
D21S1
D21S11
limit of chromosome 21 region
D110G6
pRT1 - 2.1
pRT1 - 1.1
CII49/39
?
λEag118 - 1.0


A151E3
D21S111
D21S12
D21S99
A228D8
D21S93


REGION II
D142H8
GART
SON3
C14B2
D21S65
IIO8


REGION III
A222A12
D21S17
B134B9
D21S55
A125B12
D21S60
ERG
B19C12
ERG

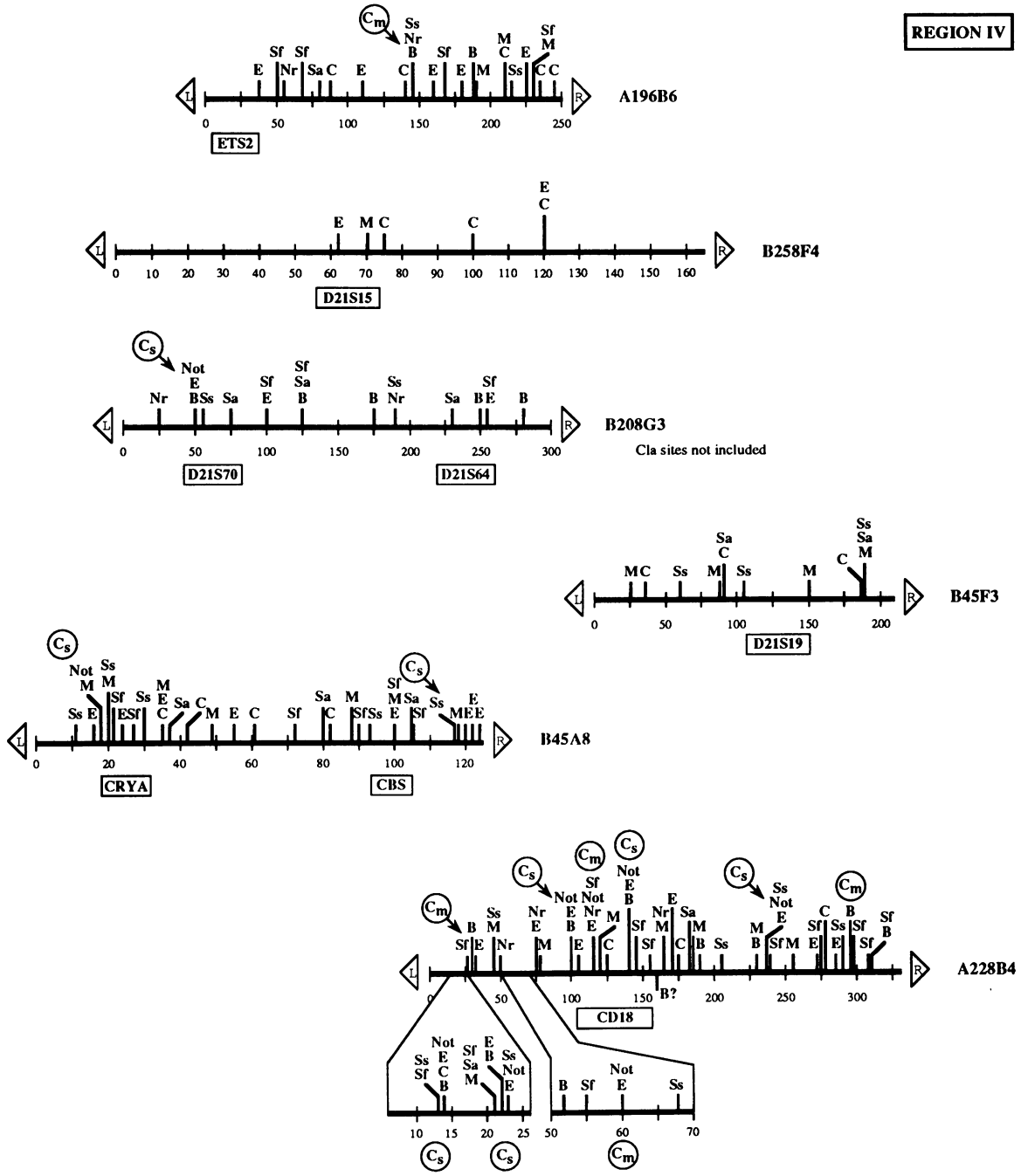


Figure 3 Restriction maps of the YACs. Pulsed-field restriction maps constructed using partial and complete digests with nine enzymes are shown, in order from the centromere, by region. The library address for each YAC is indicated (for the probe used in screening the library, see table 3). All probes known within each clone are shown boxed at their approximate positions. Size markers are in kilobases. Note that the scale varies among the maps. N = *Not*I; B = *Bss*HII; M = *Mlu*I; Nr = *Nru*I; E = *Eag*I; Ss = *Sst*II; C = *Cla*I; Sa = *Sal*I; Sf = *Sfi*I; and Cs and Cm = strong and moderate CpG clusters, respectively. Left (L) and right (R) ends with respect to the pYAC4 vector sequences are indicated. The map of D142H8 is from Gnirke et al. (1991), with *Eag*I and *Sst*II sites added. In chimeric or potentially chimeric YACs, the region confirmed from chromosome 21 is indicated.

FISH. However, the additional hybridization is seen near the centromere of either chromosome 14 or chromosome 15 and therefore may be due to pericentromeric sequence homology as reported by Van Camp et al. (1992). In region IV, the situation with yD21S15 is similar, showing hybridization to 21q22 (where D21S15 maps) but also to 21qcen. (2) yD21S11 appears to be chimeric by FISH, with sequences also mapping to chromosome 6 or chromosome 7, near the centromere. Vectorette PCR and left-end rescue were unsuccessful in generating useful end probes from this YAC. Although a triply chimeric YAC is not suggested here by FISH, for further analyses we used the conservative estimate that only ~200 kb surrounding D21S11 and D21S1 is from chromosome 21. (3) The map of yAPP provides information on the size of the APP gene. Previously APP had only been determined to be >50 kb in size (Lemaire et al. 1989), because not all introns had been cloned in their entirety. In *Eag*I digests of yAPP D110G6, probes for the promoter, 5′ and 3′ cDNA fragments together detect five fragments (data not shown). Comparison of these data—and of similar data, which resulted from use of *Sal*I, *Bss*HII, and *Sfi*I—with the map in figure 3 implies that the entire gene minimally spans ≥300 kb. When all nonduplicated chromosome 21 material is considered, YACs from the AT-rich, gene-poor region I contain 2,870 kb of DNA; those from the GC-rich, gene-rich regions II and IV contain 1,050 and 1,375 kb, respectively; and those from the AT-rich, gene-rich region III contain 1,010 kb.

Implications for Methylation

The only DNA modification seen in mammalian DNA is 5′Me-C, occurring largely in CpG dinucleotides (Bird 1986). Currently, too few unique sequences are available from these YACs to directly determine the methylation status of the majority of rare sites shown in the maps of figure 3 (for methylation of CpG islands, see below). However, some indications of the frequency of methylation in the different regions can be obtained by comparing average sizes of fragments seen in the YACs with those obtained from analysis of human lymphocyte DNA. We used the maps in figure 3 to calculate the number and size of the fragments in each YAC seen with seven enzymes (*Cla*I and *Sal*I were excluded). Data were pooled for YACs within a single region, because they share similar base composition and rare-site frequencies (Gardiner 1990; Gardiner et al. 1990*a*, 1990*b*). In regions I and III, a more realistic picture was obtained by also considering end fragments, because the number of internal fragments was often small (because of the low frequency of rare sites). A similar inclusion did not affect data in regions II and IV. Genomic pulsed-field data (for human lymphocyte DNA) were obtained from previously published physical maps (Gardiner 1990; Gardiner et al. 1990*b*) and include information for additional probes mapping to the same regions as these YACs.

Results of these comparisons are shown in table 3. In regions I and III, most enzymes give fragments 10–20 times larger in genomic DNAs, implying that 90%–95% of the rare sites are methylated. In regions II and IV, this remains true for *Mlu*I and *Nru*I only. For the sites most frequently seen in CpG islands (*Not*I, *Bss*HII, *Eag*I, and *Sst*II) (Bird 1989), however, fragments are not so much larger. They exceed those seen in YACs only by a factor of 3–6, implying that, for these enzymes, on the order of 65%–80% of sites are methylated. *Sfi*I sizes are roughly the same in the YACs and in genomic DNA, as expected for a site that is largely lacking in CpG. YACs from regions I and III, however, contain three or four times as many *Sfi*I sites as expected, suggesting some methylation. This is consistent with the observation of Butler et al. (1992) for the region around D21S13 and D21S16. In summary, not only are sites for most rare-cutters less frequent in regions I and III, but they are also less frequently unmethylated.

Potential CpG Islands

Scanning the YACs for potential CpG islands—and thus the 5′ ends of genes (Bird 1987)—may give insights into variations in gene density. Bird (1989) has shown that four of the enzymes used in this analysis (*Not*I, *Bss*HII, *Eag*I, and *Sst*II) are statistically more frequently found in CpG islands, while others, such as *Mlu*I and *Nru*I, are more often found outside CpG islands. In characterizing potential islands, we defined a strong cluster of sites as one containing three or more coincident "island" sites and defined a moderate cluster as one containing two.

Inspection of the maps in figure 3 shows that four YACs (GART, D21S65, CBS, and CD18) stand out as having both large numbers of island sites and large numbers of island-site clusters. This is in general not surprising, because previous studies indicated that these probes map to the gene-rich region of the chromosome and that D21S65, CBS, and CD18 also map to GC-rich regions (Gardiner 1990; Gardiner et al.

Table 3

Implications for Methylation

	Size of Fragment[a] (kb)							
	Region I		Region III		Region II		Region IV	
	Y	G	Y	G	Y	G	Y	G
*Not*I	>320	2,000	>336	1,300	210	330	76	550
*Bss*HII ...	113	1,750	108	800	40	120	55	300
*Eag*I	100	800	115	800	45	110	30	80
*Sst*II	100	1,000	75	800	65	150	40	250
*Mlu*I	160	2,000	100	1,900	140	1,300	35	700
*Nru*I	170	2,460	80	1,500	100	1,500	100	650
*Sfi*I	110	480	70	200	90	90	45	100

Note.—YACs were classified as belonging to one of four regions as indicated in fig. 1.

[a] Y = average sizes observed in YAC clones of internal and end fragments calculated by inspection of the maps in fig. 3; and G = average sizes of fragments observed in human genomic DNA for probes mapped to these regions (from Gardiner et al. 1990*b*). Note that I and III are the AT-rich regions; these YACs contain a single *Not*I site. The most striking regional differences are for the CpG island sites: *Not*I, *Bss*HII, *Eag*I, and *Sst*II.

1990*a*, 1990*b*). Together, these YACs comprise 1,500 kb, with 26 island clusters, i.e., an average of 1 cluster/60 kb.

Seven other YACs (D21S120, D21S110, D21S12, D21S93, D21S17, D21S15, and D21S19) are in stark contrast. They also total 1,500 kb of DNA, but they contain no island clusters. Probes within four of these YACs map to 21cen through proximal 21q22.1, and D21S17 maps within 21q22.2—both AT-rich regions. D21S19 is anomalous; it is located in 21q22.3, a very gene- and GC-rich region. The remaining YACs (D21S13, D21S52, D21S11, APP, D21S55, ERG, ETS2, and D21S64) together contain >3,000 kb and typically show low frequencies of island sites, together adding only 11 more clusters.

Are Island-Site Clusters Methylated?

Strictly speaking, rare-site clusters constitute a CpG island only if they are unmethylated. We can directly assess the methylation status of some of the observed clusters by comparing fragment size in the YAC with fragment size previously observed in human lymphocyte DNA.

Previous physical mapping data indicate that the cluster in yD21S13, the left-end strong cluster in yD21S65 (Gardiner et al. 1990*b*; Gao et al. 1991*a*), the *Not*I sites and two moderate clusters between them in yGART (Lutfalla et al. 1992; S. Cheng and K. Gardiner, unpublished data), and the left-end strong cluster in yAPP (Brahe et al. 1990; F. Tassone, unpublished data) are all unmethylated. In contrast, the cluster in yD21S52, clusters in yD21S65 at 60 and 110 kb, and those in yERG, yD21S55, and yETS2 are all methylated (Gardiner et al. 1990*b*; F. Tassone, unpublished data), because of the large fragments seen with these probes in lymphocyte DNA.

How Does the Cluster Frequency Relate to Gene Frequency?

Fourteen genes are contained within 10 YACs (for anonymous transcribed sequences, see Davidson et al. 1985; Neve et al. 1987; Stefani et al. 1988). Table 4 specifies whether a CpG island association is observed for each of the 14 genes and also gives the cluster density within each corresponding YAC. Cluster density was calculated on the basis of 300 kb because this approaches the average size for these YACs. Zero, low, and high cluster density was assigned by inspection of the maps in figure 3.

Forty-five percent (6/14) of the genes are within the four YACs (1,500 kb) that show high cluster densities averaging one cluster/60 kb. Twenty percent (3/14) of the genes are located in areas with much lower cluster densities averaging 1/300 kb. A significant 35% (5/14) are in regions with no clusters observed within 300 kb. Because most housekeeping and many tissue-specific genes are associated with CpG islands (Bird 1987; Gardiner-Garden and Frommer 1987), it is possible that this analysis is actually reflecting differences in gene size and that in many cases the attendant CpG islands are merely outside the YACs

Table 4

Gene and CpG Cluster Associations

Gene[a]	CpG Island[b]	Cluster Density[c] (no./300 kb)
D21S13	Yes	Low (1)
APP	Yes	Low (1)
D21S93	No	Zero (0)
GART	Yes	High (4)
SON (GART)	Yes	High (4)
AML1[d] (S65)	Yes	High (5)
D21S17	No	Zero (0)
D21S55	No	Low (1)
D21S60 (ERG)	No	Low (1)
ERG	Yes?	Low (1)
ETS2	No?	Low (1)
CBS	Yes	High (5)
CRYA (CBS)	Yes	High (5)
CD18	Yes	High (9)

[a] From YAC studied (in parentheses is given the name of the YAC, if different from that of the gene).

[b] Yes = putative CpG island is present near the gene; and No = no apparent CpG island is in the vicinity of the gene. Status is determined by inspection of maps in fig. 3.

[c] Determined by inspection of the maps in fig. 3.

[d] Sources: Gao et al. (1991*a*) and Myoshi et al. (1991).

analyzed. Perhaps, then, apparently gene-rich regions are merely home to smaller genes. Alternatively, the data may be indicating a significant collection of genes lacking CpG islands or possessing more subtle CpG islands that are not detected by using these rare-cutting enzymes.

Discussion

We have characterized 22 YAC clones distributed throughout the long arm of chromosome 21. YACs were selected on the basis of their physical map position and are biased toward larger, nonchimeric YACs (average size 350 kb; 17/22 are nonchimeric). Only four YACs are confirmed chimeric, but each contains significant amounts of chromosome 21 material. In total, the collection contains ~6,000 kb of chromosome 21 DNA. All clones appeared stable during prolonged growth, and no trivial sequence rearrangements were observed. We can therefore conclude that these YACs provide reasonable starting material for the examination of sequence organizational features and for undertaking comprehensive gene searches.

Pulsed-field restriction maps of all 22 clones were used to investigate probe density, methylation patterns in human lymphocyte DNA, frequency of potential CpG islands, and the correlation between gene density and CpG island frequency. Nine YACs contained a total of 12 additional probes already placed on the physical map. Eight of these probes were found in five YACs mapping to region I. One YAC contains four probes within <275 kb, a second contains three within <150 kb, etc. All together, 13 probes of the 36 placed on various physical maps (Cox et al. 1990; Gardiner et al. 1990*b*; Owen et al. 1990) are clustered in ~1 Mbp of the >20 Mbp covered by the maps spanning 21cen through 21q21. It appears that the proximal two-thirds of the chromosome has been largely refractory to cloning, save for subregions from which most probes have been obtained. Reasons for this anomaly may be understood after further sequence analysis of additional YACs from this region. In contrast, in spite of the greater probe density in the distal one-third, very few probes are found clustered in the >3,000 kb of DNA analyzed.

As expected, all YACs contained many more rare sites than observed in genomic DNA, but the discrepancy was again greater both for YACs and probes within the proximal two-thirds and also for those mapping to 21q22.2. Methylation in these regions appears to be 90%–95% of rare sites, while in the remaining regions it is 65%–80%. This is inversely related to the gene density in gross terms and therefore conceivably reflects the difference in (unmethylated) CpG islands found in the latter regions.

When potential CpG island densities were considered, YACs were divided into three groups—those with clusters every 60 kb, those with clusters every 300 kb, and those with clusters spaced at >300 kb. While all YACs with high cluster densities map in the gene-rich region, the converse is not true. Roughly 1,500 kb of YAC DNA (half that analyzed from the distal one-third) showed few or no clusters, in spite of being within the gene-rich region. Furthermore, of the 14 genes mapped to these YACs, 8 map in YACs showing clusters ≥300 kb apart. This group of eight includes five genes within the distal one-third of the chromosome.

These observations lead to an interesting speculation regarding gene size and gene density. It remains perfectly possible that many genes on chromosome 21 do not have CpG islands as detected by rare restriction sites. Alternatively, the average gene size may vary considerably by region. For example, the APP gene is ≥300 kb in size; the GART gene is 40 kb (Gnirke et al. 1991). Both genes possess CpG islands, but APP

resides in a region where they are, on average, 300 kb apart, GART where they are 60 kb apart.

All YACs analyzed here are potentially relevant to chromosome 21–associated diseases. In particular, those mapping to regions II–IV are likely to contain genes affecting various aspects of the DS phenotype (McCormick et al. 1989; Korenberg 1991; Korenberg et al. 1992). Genes within yCD18 may be good candidates for HP, those within yGART or yD21S93 for ALS, those within yD21S19 for PME, and those centromeric for FAD (St George-Hyslop et al. 1987; Estabrooks et al. 1990; Lehesjoki et al. 1991; Siddique et al. 1991). Rare restriction maps of these YACs, their CpG island distributions, and methylation patterns will be of assistance in the isolation and mapping of such genes, and nonchimeric YACs will form the basis for walks through the regions in searches for additional genes.

Acknowledgments

The authors are grateful to the participants of the International Joint YAC Screening Effort who have made the rapid screening of the library and wide availability of the YACs possible. In particular, we wish to thank David Patterson, who heads the screening effort, and the following who submitted one or more primer sets used to identify the YACs discussed here: Miles Brennan, Nicole Creau-Goldberg, Jan Kraus, David Kurnit, David Patterson, Roger Reeves, Sue Rider, Gordon Stewart, and Rudy Tanzi. The authors are also grateful to Buddy Brownstein and Maynard Olson for making the YAC library available to the Eleanor Roosevelt Institute and are grateful to R. Tanzi for the gift of the unpublished APP complete cDNA clone. We also thank Diana Smith for general technical assistance and thank Iris Hart, Wen-Lin Kuo, Joe Gray, and J. G. Dauwerse for protocols and advice on FISH. This is contribution 1213 from the Eleanor Roosevelt Institute for Cancer Research. This work was supported by National Institutes of Health grants HD17449 and HG00001.

References

Berdichevskii FB, Chumakov IM, Kiselev LL (1988) Determination of the nucleotide sequence of the son3 fragment of the human genome: identification of a new protein with an unusual structure and homology with DNA-binding proteins (in Russian). Mol Biol (Mosk) 22:794–801

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

——— (1987) CpG islands as gene markers in the vertebrate nucleus. Trends Genet 3:342–347

——— (1989) Two classes of observed frequency for rare-cutter sites in CpG islands. Nucleic Acids Res 17:9485

Brahe C, Tassone F, Millington-Ward A, Serro A, Gardiner K (1990) Potential gene sequence isolation and regional mapping in human chromosome 21. Am J Med Genet 7: 120–124

Brownstein BH, Silverman GA, Little RD, Burke DR, Korsmeyer SJ, Schlessinger D, Olson MV (1989) Isolation of single-copy human genes from a library of yeast artificial chromosome clones. Science 244:1348–1351

Burke DT, Carle GF, Olson MV (1987) Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. Science 236:806–811

Burmeister M, Kim S, Price ER, de Lange T, Tantravahi U, Myers RM, Cox DR (1991) A map of the distal region of the long arm of human chromosome 21 constructed by radiation hybrid mapping and pulsed-field gel electrophoresis. Genomics 9:19–30

Butler R, Ogilvie DJ, Elvin P, Riley JH, Finniear RS, Slynn G, Morten JEN, et al (1992) Walking, cloning, and mapping with yeast artificial chromosomes: a contig encompassing D21S13 and D21S16. Genomics 12:42–51

Cox DR, Burmeister M, Price ER, Kim S, Meyer RM (1990) Radiation hybrid mapping: a somatic cell genetic method for constructing high-resolution maps of mammalian chromosomes. Science 250:245–250

Davidson JN, Rumbsy G, Niswander LA (1985) Expression of genes on human chromosome 21. Ann NY Acad Sci 450:43–54

Epstein CJ (1986) The consequences of chromosome in balance: principles, mechanisms, and models. Cambridge University Press, New York

Estabrooks LL, Rao KW, Donahue RP, Aylsworth AS (1990) Holoprosencephaly in an infant with a minute deletion of chromosome 21(q22.3). Am J Med Genet 36: 306–309

Feinberg AP, Vogelstein B (1989) Addendum to "A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity." Anal Biochem 137: 266–267

Gao J, Erickson P, Gardiner K, LeBeau M, Diaz M, Patterson D, Rowley J, et al (1991*a*) Isolation of a yeast artificial chromosome spanning the 8;21 translocation breakpoint t(8;21)(q22;q22.3) in acute myelogenous leukemia. Proc Natl Acad Sci USA 88:4882–4886

Gao J, Erickson P, Patterson D, Jones C, Drabkin H (1991*b*) Isolation and regional mapping of *Notl* and *Eagl* clones from human chromosome 21. Genomics 10:166–172

Gardiner K (1990) Physical mapping of the long arm of chromosome 21. In: Patterson D, Epstein CJ (eds) Molecular genetics of chromosome 21 and Down syndrome. Wiley-Liss, New York, pp 1–14

Gardiner K, Aissani B, Bernardi G (1990*a*) A compositional map of human chromosome 21. EMBO J 9:1853–1858

Gardiner K, Horisberger M, Kraus J, Tantravahi U, Koren-

berg J, Rao V, Reddy S, et al (1990*b*) Analysis of human chromosome 21: correlation of physical and cytogenetic maps; gene and CpG island distributions. EMBO J 9: 25–34

Gardiner-Garden M, Frommer M (1987) CpG islands in vertebrate genomes. Mol Biol 196:261–282

Gnirke A, Barnes TS, Patterson D, Schild D, Featherstone T, Olson MV (1991) Cloning and *in vivo* expression of the human GART gene using yeast artificial chromosomes. EMBO J 10:1629–1634

Goate A, Chartier-Harlin MC, Mullan M, Brown J, Crawford F, Fidani L, Guiffra L, et al (1991) Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease. Nature 349:704–706

Green ED, Olson MV (1990) Systematic screening of yeast artificial-chromosome libraries by use of the polymerase chain reaction. Proc Natl Acad Sci USA 87:1213–1217

Imai T, Olson MV (1990) Second-generation approach to the construction of yeast artificial-chromosome libraries. Genomics 8:297–303

Korenberg JR (1991) Down syndrome phenotypic mapping. In: Epstein CJ (ed) The morphogenesis of Down syndrome. Wiley-Liss, New York, pp 43–52

Korenberg JR, Bradley C, Disteche CM (1992) Down syndrome: molecular mapping of the congenital heart disease and duodenal stenosis. Am J Hum Genet 50:294–302

Korenberg JR, Kawashima H, Pulst S-M, Ikeuchi T, Ogasawara N, Yamamoto K, Schonberg SA, 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

Lehesjoki A-E, Koskiniemi M, Sistonen P, Miao J, Hastbacka J, Norio R, de la Chapelle A (1991) Localization of a gene for progressive myoclonus epilepsy to chromosome 21q22. Proc Natl Acad Sci USA 88:3696–3699

Lehrach H, Drmanac R, Hoheisel J, Larin Z, Lennon G, Monaco AP, Nizetic D, et al (1990) Hybridization fingerprinting in genome mapping and sequencing. In: Davies KE, Tilghman SM (eds) Genetic and physical mapping. Vol. 1: Genome analysis. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, pp 39–81

Lemaire HG, Salbaum JM, Multhaup G, Kang J, Bayney RM, Unterbeck A, Beyreuther K, et al (1989) The $PreA4_{695}$ precursor protein of Alzheimer's disease A4 amyloid is encoded by 16 exons. Nucleic Acids Res 17:517–522

Lutfalla G, Gardiner K, Proudhon D, Vielh E, Uze G (1992) The structure of the human interferon alpha/beta receptor gene. J Biol Chem 267:2802–2809

McCormick MK, Schinzel A, Petersen MB, Stetten G, Driscoll DJ, Cantu ES, Tranebjaerg L, et al (1989) Molecular genetic approach to the characterization of the "Down syndrome region" of chromosome 21. Genomics 5:325–331

Mendez MJ, Klapholz S, Brownstein BH, Gemmill RM (1991) Rapid screening of a YAC library by pulsed-field gel Southern blot analysis of pooled YAC clones. Genomics 10:661–665

Miyoshi H, Shimizu K, Kozu T, Masheki N, Kaneko Y, Ohki M (1991) t(8;21) breakpoints on chromosome 21 in acute myeloid leukemia are clustered within a limited region of a single gene 1AML1. Proc Natl Acad Sci USA 88:10431–10434

Nelson DL, Ledbetter SA, Corbo L, Victoria MF, Ramirez-Solis R, Webster TD, Ledbetter DH, et al (1989) *Alu* polymerase chain reaction: a method for rapid isolation of human-specific sequences from complex DNA sources. Proc Natl Acad Sci USA 86:6686–6690

Neve RL, Stewart GD, Newcomb P, Van Keuren ML, Patterson D, Drabkin HA, Kurnit DM (1987) Chromosome 21–encoded cDNA clones. Gene 49:361–369

Owen MJ, James LA, Hardy JA, Williamson R, Goate AM (1990) Physical mapping around the Alzheimer disease locus on the proximal long arm of chromosome 21. Am J Hum Genet 46:316–322

Patterson D (1991) Report of the Second International Workshop on Human Chromosome 21 Mapping. Cytogenet Cell Genet 57:167–174

Pueschel SM (1982) Biomedical aspects in Down syndrome. In: Pueschel SN, Rynders JF (eds) Down syndrome: advances in biomedicine and the behavioral sciences. Ware, Cambridge, MA

Rahmani Z, Blouin J-L, Creau-Goldberg N, Watkins PC, Mattei J-F, Poissonnier M, Prieur M, et al (1989) Critical role of the *D21S55* region on chromosome 21 in the pathogenesis of Down syndrome. Proc Natl Acad Sci USA 86:5958–5962

Riley J, Butler R, Ogilvie D, Finniear R, Jenner D, Powell S, Anand R, et al (1990) A novel, rapid method for the isolation of terminal sequences from yeast artificial chromosome (YAC) clones. Nucleic Acids Res 18:2887–2890

St George-Hyslop PH, Tanzi RE, Polinsky RJ, Haines JL, Nee L, Watkins PC, Myers RH, et al (1987) The genetic defect causing familial Alzheimer's disease maps on chromosome 21. Science 235:885–890

Salbaum JM, Weidemann A, Lemaire H-G, Masters CL, Beyreuther K (1988) The promoter of Alzheimer's disease amyloid A4 precursor gene. EMBO J 7:2807–2813

Schellenberg GD, Anderson L, O'dahl S, Wisjman EM, Sadovnick AD, Ball MJ, Larson EB, et al (1991) APP_{717}, APP_{693}, and PRIP gene mutations are rare in Alzheimer disease. Am J Hum Genet 49:511–517

Schwartz DC, Smith LC, Baker M, Hsu M (1989) ED: pulse field electrophoresis instrument. Nature 342:575–576

Siddique T, Figlewicz DA, Pericak-Vance MA, Haines JL, Rouleau G, Jeffers AJ, Sapp P, et al (1991) Linkage of a gene causing familial amyotrophic lateral sclerosis to chromosome 21 and evidence of genetic-locus heterogeneity. N Engl J Med 324:1381–1384

Stefani L, Galt J, Palmer A, Affara N, Ferguson-Smith M,

Nevin NC (1988) Expression of chromosome 21 specific sequences in normal and Down's syndrome tissues. Nucleic Acids Res 16:2885–2895

Tanzi RE, Gusella JF, Watkins PC, Bruns GAP, St George-Hyslop P, Van Keuren ML, Patterson D, et al (1987) The amyloid b protein gene: cDNA cloning, mRNA distribution, and genetic linkage near the Alzheimer locus. Science 235:880–884

Trask B, Pinkel D, Van Den Engh G (1989) The proximity of DNA sequences in interphase cell nuclei is correlated to genomic distance and permits ordering of cosmids spanning 250 kilobase pairs. Genomics 5:710–717

Van Camp G, Cruts M, Backhovens H, Wehnert A, Van Broeckhoven C (1992) Unique sequence homology in the pericentromeric regions of the long arms of chromosomes 13 and 21. Genomics 12:158–160