A survey of the genomic distribution of alpha satellite DNA on all the human chromosomes, and derivation of a new consensus sequence

K.H.Choo*, B.Vissel, A.Nagy, E.Earle and P.Kalitsis
Murdoch Institute for Research into Birth Defects, Royal Children's Hospital, Flemington Road, Parkville 3052 Victoria, Australia

Received December 18, 1990; Revised and Accepted February 25, 1991

INTRODUCTION

The human centromere is characterised by a major class of repetitive DNA known as the alpha satellite DNA (Manuelidis, 1978; Mitchell et al 1985). This DNA makes up 3–5% of each chromosome and has a monomeric repeating unit of 171 bp (Singer, 1982; Vissel and Choo, 1987; Waye and Willard, 1987). These monomeric units are organised into different tandem arrays that constitute clearly definable higher-order repeating structures or alphoid subfamilies. To date, at least 33 different alphoid subfamilies have been identified. Some of these subfamilies are specific for a single chromosome, while others are common to a small group of chromosomes. In addition, some chromosomes appear to have only a single subfamily within their centromeres, whereas other chromosomes may possess several different subfamilies.

In this communication, we have surveyed all the alpha satellite subfamilies that have been reported. These subfamilies are systemmatically listed according to their chromosomal specificities and their multiplicity of sharing. Based on an analysis of 293 individual 171-bp monomeric units originating from all the 24 different human chromosomes, we have derived an updated consensus sequence for the human alpha satellite DNA. The information presented should be useful for the understanding of the structural organisation of the human centromeres, and the process by which homologous and non-homologous chromosomes interact in the evolution and maintanence of the different subfamilies (Choo, 1990). As the chromosome-specific alphoid sequences are increasingly being used as a tool for chromosomal studies and for centromere-based linkage analysis, the present survey provides a reference to assist in the selection of the appropriate sequences for such purposes.

Chromosomal distribution of alpha satellite DNA

Table 1 summarises the properties and chromosomal distribution of all the known human alpha satellite subfamilies. In this Table, the subfamilies are classified according to whether they are specific for a single chromosome (Group A), or whether they are shared by more than one chromosome (Group B). Three groups of chromosomes show the sharing property: (a) chromosomes 1, 5 and 19; (b) chromosomes 4 and 9; and (c) chromosomes 13, 14, 21 and 22. As can be seen from the Table, despite this sharing, a different set of subfamilies specific for each of chromosomes 1, 4, 9, 14 and 22 has also been found. Thus, except for chromosomes 5 and 19, and chromosomes 13 and 21 (see below), a chromosome-specific alpha subfamily has been identified on all the human chromosomes. Table 1 also provides information on the size of the cloned sequences, the size of their corresponding genomic higher-order structures, the enzymes that define these higher-order structures, and the number of monomers that have been sequenced and used for the derivation of the new consensus sequence.

Table 2 regroups the alphoid subfamilies shown in Table 1 according to their presence or co-existence within a single chromosome. While several of the chromosomes contain only a single alphoid subfamily, the remaining chromosomes have more than one subfamily. With chromosomes 13, 14 and 21, as many as five to seven structurally distinct subfamilies have so far been characterised. It is likely that the number of different alphoid subfamilies on at least some of the chromosomes will increase as new subfamilies become identified.

Most of the sequences listed in Tables 1 and 2 are useful as probes for *in situ* hybridisation studies for the identification of specific human metaphase chromosomes relating to translocations, marker chromosomes and aneuploidy. Two groups of chromosomes, 13 and 21, and 5 and 19, cannot be separately detected by the known alphoid probes (alphaRI or L1.26, and pG-A16, respectively). Similarly, chromosome 14 can only be detected using a probe (alpha-XT) that detects both 14 and 22, although an additional probe (p22/1:2.1) allows the separate identification of chromosome 22. Probes p82H, pTRA-1, -4, and -7 are not suitable for *in situ* hybridisation experiments because of their relatively low genomic copy number. Due to the highly polymorphic nature of alpha satellite DNA, most of the probes listed should be useful in centromere-based linkage analysis of the different chromosomes. Warren *et al* (1990) and Jabs *et al* (1991) have recently reported alpha satellite polymorphisms that are distinct for chromosomes 21 and 13, respectively.

Consensus sequence of human alpha satellite DNA

This analysis is an extension of the data of two previous studies (Vissel and Choo, 1987; Waye and Willard, 1987). The sources

* To whom correspondence should be addressed

Table 1. Properties and chromosomal distribution of human alpha satellite DNA. In this Table, only the earliest reference describing a new subfamily is listed except: (i) when two or more references appeared closely; (ii) when a subsequent reference adds new information relevant to the Table (such as sequence data or definition of higher-order structures); and (iii) when the sizes of the higher-order units appear to be slightly different and it is not clear if these belong to the same subfamily (eg. see chromosomes 3 and Y). n.d. = not defined. 'SIZE' refers to the size of the cloned insert. The number of sequenced, complete monomers used for the derivation of the consensus sequence (see Fig.1) is shown under the 'SEQUENCE' column. In addition to the monomers listed in the Table, the following cloned monomers of unknown chromosomal origin are also used in the consensus sequence analysis: 8 monomers from XI(1020)22-73 and -82 (Jorgensen *et al*, 1987); 4 monomers from pE1 (Gray *et al*, 1985); and 5 monomers from pC21 (Jones and Potter, 1985) (also refer to Waye and Willard, 1987). Ac = AccI, Ba = BamHI, Bg = BglII, Bs = BstNI, Dr = DraI, Ec = EcoRI, EcV = EcoRV, Ha = HaeIII, HdII = HindII, HdIII = HindIII, Hf = HinfI, Ms = MspI, Ns = NsiI, Ps = PstI, Pv = PvuII, Rs = RsaI, Sa = Sau3A, Sp = SphI, Ss = SstI, St = StuI, Tq = TaqI and Xb = XbaI.

CHROMOSOME		INFORMATION ABOUT CLONE			SOUTHERN DATA		REFERENCE
	note	DESIGNATION	SIZE	SEQUENCE	HIGHER ORDER STRUCTURE	ENZYME	
(A) Chromosome -Specific							
1	a	pSD1-1, -2	1.9 kb	11-mer	1.9 kb (11-mer)	HdIII,Xb,Sp	Waye et al., 1987b; Willard and Waye, 1987
		(see B, below)					
2		pBS4D	1.4 kb	8-mer	0.68 kb (4-mer)	Hf, Xb, HdIII	Rocchi et al., 1990
3	b	p3-9	2.4 kb	6-mer	2.9 kb (17-mer)	Ac,Dr,HdIII,Ps,Pv,Sa	Waye & Willard, 1989
	b	VIIB4	0.64 kb	3-mer	2.75 kb (16-mer)	HdIII, Tq	Dellatre et al.,1988
4	c	(see B, below)					
5		(see B, below)					
6	d	p308	2.86 kb	14-mer	2.86 kb (17-mer)	Ba, Tq	Jabs et al., 1984; Jabs and Perisco, 1987
7		pMGB7	2.7 kb	16-mer	2.7 kb (16-mer)	HdIII,Ha,Tq	Waye et al., 1987c
	e	palpha7-d1 and -t1	0.34 and 0.68 kb	2-&4-mer	0.34+0.68 kb (6-mer)	Ec	Waye et al., 1987c
8		pJM128	n.d.	9-mer	n.d.	n.d.	A. Wyman, unpubl. (in Waye and Willard, 1987)
9		pMR9A	0.34 kb	2-mer	2.7 kb (16-mer)	Ps,Rs,Ms	Rocchi et al, 1991
		(see B, below)					
10		pALPHA10RP8 & RR6	0.95 and 1.02 kb	n.d.	1.35 kb (8-mer)	Rs,Ac,Ps,Sa,St,Ha	Devilee et al., 1988
11	a	pLC11-A and -B	0.85 kb	5-mer	0.85 kb (5-mer)	Xb	Waye et al., 1987a
12		pBR12	0.68 kb	3-mer	1.4 kb (8-mer)	HdIII,Tq,Pv	Baldini et al., 1990; Embl library
		pAlpha12H8	1.4 kb	n.d.	1.4 kb (8-mer)	HdIII, Pv	Looijega et al., 1990
13		(see B, below)					
14	f	p82H	2.4 kb	2-mer	n.d.	n.d.	Mitchell et al., 1985; Waye et al., 1988
	g	pTRA-54	n.d.	n.d.	n.d.	n.d.	Vissel and Choo, 1991
		(see B, below)					
15	h	pTRA-20	1.8 kb	n.d.	2.5 kb (14-mer)	Rs,Dr	Choo et al., 1990
	h	pTRA-25	4.5 kb	n.d.	4.5 kb (26-mer)	Ac,Bg,Bs,HdII,Ns	Choo et al., 1990
16		pSE16	0.68 kb	4-mer	1.7 kb (10-mer)	Sa, EcV	Greig et al., 1989
17	i	p17H8	2.7 kb	16-mer	2.05-2.74kb (12 to 16-mer)	Ec,Pv	Waye and Willard, 1986
	i	TR17	1.6 kb	n.d.	2.05-2.74kb (12 to 16-mer)	Ps	Choo et al., 1987
18		L1.84	0.68 kb	4-mer	0.68 kb (4-mer)	Ec	Devilee et al., 1986a; 1986b
19		(see B, below)					
20		p3-4	n.d.	8-mer	n.d.	n.d.	J. Waye, unpubl (in Waye and Willard, 1987)
21		(see B, below)					
22	a	p22/1:2.1;2.8;0.73	2.1 kb	2-mer	2.1 and 2.8 kb	Ec	McDermid et al., 1986
		(see B, below)					
X		pBamX7	2.0 kb	12-mer	2.0 kb (12-mer)	Ps, Ba, Ss	Willard et al., 1983; Waye and Willard, 1985
		pXBR-1	2.0 kb	n.d.	2.0 kb (12-mer)	Ps, Ba	Yang et al., 1982
Y	a,j	Y84 and Y97	16.5 and 5.5 kb	7-mer	5.5 kb (32-mer)	Ec,Ps,HdIII	Wolfe et al., 1985
	k,j	cY73	Approx. 37 kb	12-mer	5.7 kb (33-mer)	Ec,Ps,HdIII	Tyler-Smith and Brown, 1987
					and 6.0 kb (35-mer)	Ec, Ps	Tyler-Smith and Brown, 1987
(B) Present on multiple chromosomes							
1, 5 ,19		pC1.8	17 kb	8-mer	n.d.	n.d.	Baldini et al., 1989
4, 9		pG-XbaII/340	0.34 kb	2-mer	1.2 kb (7-mer)	Ps, Rs, Tq	Hulsebos et al., 1988
5, 19		pG-A16	1.95 kb	2-mer	2.25 kb (13-mer)	Ms, Ps, Rs, Ss,Ec	Hulsebos et al., 1988
13, 21	l	AlphaR1 and L1.26	0.68 and 0.85 kb	4-&5-mer	0.68 kb (4-mer)	Ec,Xb	Jorgensen et al., 1987; Devilee et al., 1986a; 1986b
14, 22	m	AlphaXT(680),14-12	0.68 kb	8-mer	1.36 kb (8-mer)	Xb, Tq	Jorgensen et al., 1988
13, 14, 21		pTRA-2	3.95 kb	22-mer	3.95 kb (23-mer)	Hi,Rs,Ms	Choo et al., 1988; Vissel and Choo, 1991
13, 14, 21		pTRA-1	1.20 kb	7-mer	n.d.	n.d.	Choo et al., 1989; Vissel and Choo, 1991
13, 14, 21		pTRA-4	5.08 kb	28-mer	n.d.	n.d.	Choo et al., 1989; Vissel and Choo, 1991
13, 14, 21		pTRA-7	1.71 kb	9-mer	n.d.	n.d.	Choo et al., 1989; Vissel and Choo, 1991

Note a: Where multiple clones of a subfamily were described and/or sequenced, the sequence data of only one of the clones is used in the consensus sequence analysis: pSDI-1, pLC11-A, p22/1:0.73 and Y97 for chromosomes 1, 11, 22 and Y, respectively.
Note b: Not known if the two clones represent the same subfamily.
Note c: Alphoid probes specific for this chromosome has been described, but not published (Oncor, inc; Willard, unpubl.)
Note d: The monomers in the sequence reported by Jabs and Perisco (1987) were arranged in units of 173 bp. These are reorganised into 171 bp units by realignment to achieve maximum homology with the standard consensus sequence.
Note e: In addition to the sequences of Waye *et al* (1987c), those of 31 monomers from two other chromosome 7-specific alphoid clones [RI(340)-1′ and −2′] are also included in the present consensus sequence (Jorgensen *et al*, 1986).
Note f: Mitchell *et al* (1985) reported a sequence of 10 monomeric units and derived a dimeric consensus sequence for these units. This dimeric sequence is used in the present consensus sequence analysis.
Note g: This subfamily was identified in the course of the analysis of somatic cell hybrid DNA but has as yet not been cloned out (Vissel and Choo, 1991).
Note h: In addition to these two subfamilies, there is preliminary evidence for at least one other subfamily on chromosome 15 (Choo *et al*, 1990).
Note i: Clones corresponding to the other slightly less abundant polymorphic higher-order structures (12- to 15-mers) have also been described but are not listed here (Waye and Willard, 1986; Choo *et al*, 1987).
Note j: Although the reported sizes for the predominant higher-order unit for Y84 (or Y97) and cY73 are different (5.5 kb and 5.7 kb, respectively), they most likely represent the same subfamily structure as evident from their definition by the same restriction enzymes. The discrepancy may be due to errors in size ascertainment.
Note k: Nine cosmid clones for this subfamily were described. Cosmid cY73 is listed here since its 5.7 kb and 6.0 kb higher-order units were subcloned and sequenced. The sequence revealed a deletion of two complete 171 bp monomers in the 5.7 unit, but otherwise a very low degree of sequence divergence (<0.1%) between the two structures. As such, only the sequence of the predominant 5.7 unit is used in the consensus sequence analysis.
Note l: Jorgensen *et al* (1987) described the sequences of 8 tetrameric (680 bp) clones for the alpha RI subfamily but only that for clone alpha RI (680) 21-208 is used in the consensus sequence analysis.
Note m: Multiple clones for this subfamily were described (Jorgensen *et al*, 1988). Only one [alpha-XT(680),14-12] is listed here. In addition, one representative subclone for each of the regions covering the entire 1,360 bp higher-order structure is used for the derivation of the consensus sequence. These subclones are: 340 bp each of αXT(680),14-12 and αX(340),14-12; and 680 bp of αT(1360),14-12.

Table 2. Distribution and co-existence of alphoid subfamilies on individual human chromosomes.

Chromosome	Alpha subfamily	Present on chromosomes	No.of subfamily per chromosome
1	pSD1-1, -2	1	2
	pC1.8	1, 5, 19	
2	pBS4D	2	1
3	p3-9	3	1−2*
	VIIB4	3	
4	Oncor, inc.	4	2
	pG-XbaII/340	4, 9	
5	pG-A16	5, 19	2
	pC1.8	1, 5, 19	
6	p308	6	1
7	pMGB7	7	2
	pAlpha7-d1,-t1	7	
8	pJM128	8	1
9	pMR9A	9	2
	pG-XbaII/340	4,	9
10	pAlpha10-RP8,-RR6	10	1
11	pLC11-A,-B	11	1
12	pBR12;pAlpha12H8	12	1
13	AlphaR1; L1.26	13, 21	5
	pTRA-2	13, 14, 21	
	pTRA-1	13, 14, 21	
	pTRA-4	13, 14, 21	
	pTRA-7	13, 14, 21	
14	p82H	14	7
	pTRA-54	14	
	Alpha-XT	14, 22	
	pTRA-2	13, 14, 21	
	pTRA-1	13, 14, 21	
	pTRA-4	13, 14, 21	
	pTRA-7	13, 14, 21	
15	pTRA-20	15	>2*
	pTRA-25	15	
16	pSE16	16	1
17	p17H8; TR17	17	1
18	L1.84	18	1
19	pC1.8	1, 5, 19	2
	pG-A16	5, 19	
20	p3-4	20	1
21	AlphaR1; L1.26	13, 21	5
	pTRA-2	13, 14, 21	
	pTRA-1	13, 14, 21	
	pTRA-4	13, 14, 21	
	pTRA-7	13, 14, 21	
22	p22/1:2.1;1:2.8;1:0.73	22	2
	Alpha-XT	14, 22	
X	pBamX7; pXBR-1	X	1
Y	Y84; Y97; cY73	Y	1

* see Table 1 for explanation

and the number of monomers of human alpha satellite DNA that are used for the derivation of the new consensus sequence in Figure 1 are listed in Table 1. For this analysis, only cloned, complete alphoid monomers are used. Where multiple subclones of the same subfamily or higher-order structure have been reported, only the sequence of one subclone is used. Altogether, the new consensus sequence is based on 293 individually cloned 171 bp alphoid monomers, representing 28 distinct alphoid subfamilies originating from every human chromosome. It should be noted that in the derivation of this consensus sequence, the genomic 'weighting' of the relative abundance of the different subfamilies has not been taken into account. Such 'weighting' is difficult to be carried out reliably because of (a) intrinsic inaccuracies associated with the estimation of the copy number of different subfamilies, and (b) extensive polymorphic variation in the size of alphoid arrays on different homologous chromosomes. At the most basic level, the present consensus sequence therefore represents an 'evolutionary consensus' which reveals the degree of tolerence for the conservation (or divergence) of nucleotide in the different positions within the alphoid monomers of different subfamilies. Information for such a consensus sequence is important for the understanding of the evolution of this DNA and its potential biological roles. The present consensus sequence should serve as a base for future updating as new alphoid subfamilies and sequence data become available.

ACKNOWLEDGEMENTS

We thank the National Health and Medical Research Council for support. KHC is a Senior Research Fellow of the Council.

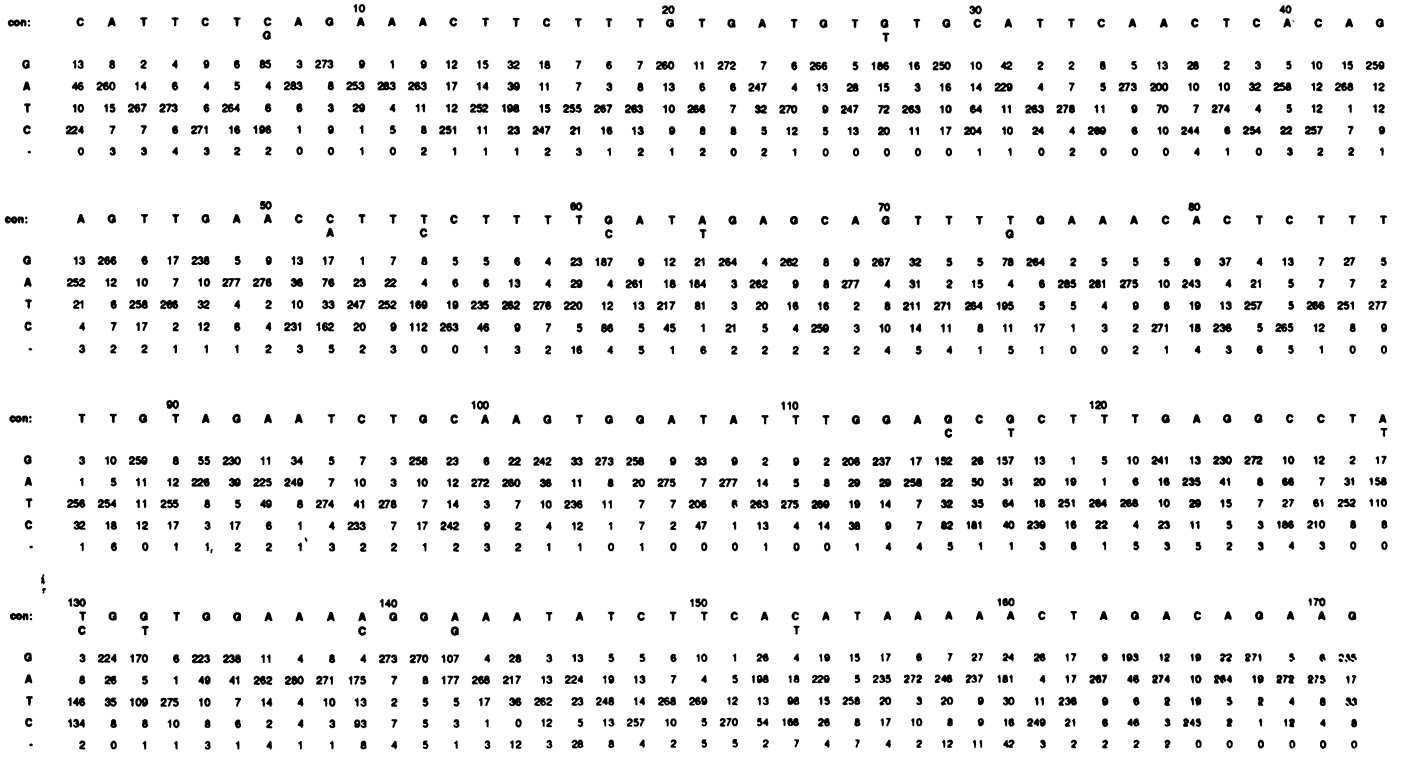

Fig. 1. Derivation of human alpha satellite monomer consensus sequence (con) from 293 cloned monomers of wide-ranging chromosomal origins (refer to Table 1). The composition of the four bases (G, A, T, C) is shown. (−) indicates a gap introduced in the monomer to allow optimal sequence alignment. Assignment of a consensus base is based on its presence at three times or more than the next abundant base; otherwise, the two most highly represented bases are used as alternative consensus bases.

REFERENCES

1. Baldini A., Rocchi M., Archidiacono N., Miller O.J. and Miller D.A. (1990) Am. J. Hum. Genet. 46: 784−788.
2. Baldini A., Smith D.I., Rocchi M., Miller O.J. and Miller D.A. (1989) Genomics 5: 822−828.
3. Choo K.H. (1990) Mol. Biol. Med., 7: 437−449.
4. Choo K.H., Brown R., Webb G., Craig I.W. and Filby R.G. (1987) DNA 6: 297−305.
5. Choo K.H., Earle E., Vissel B. and Filby R.G. (1990) Genomics 7: 143−151.
6. Choo K.H., Vissel B., Brown R., Filby R.G. and Earle E. (1988) Nucl. Acids Res. 16: 1273−1284.
7. Choo K.H., Vissel B. and Earle E. (1989) Genomics 5: 332−344
8. Delattre O., Bernard A., Malfoy B., Marlhens F., Viegas-Pequignot E., Brossard C., Haguenauer O., Creau-Goldberg N., N'Guyen Van Cong, Dutrillaux B. and Thomas G. (1988) Hum. Hered. 38: 156−167.
9. Devilee P., Kievits T., Waye J.S., Pearson P.L. and Willard H.F. (1988) Genomics 3: 1−7.
10. Devilee P., Cremer T., Slagboom P., Bakker E., Scholl H.P., Hager H.D., Stevenson A.F.G., Cornelisse C.J. and Pearson P.L. (1986a) Cytogenet. Cell Genet. 41: 193−201.
11. Devilee P., Slagboom P., Cornelisse C.J. and Pearson P.L. (1986b) Nucl. Acids Res. 14: 2059−2073.
12. Gray K.M., White J.W., Costanzi C., Gillespie D., Schroeder W.T., Calabretta B and Saunders G.F. (1985) Nucl. Acids Res. 13: 521−535
13. Greig G.M., England S.B., Bedford H.M. and Willard H.F. (1989) Am. J. Hum. Genet. 45: 862−872.
14. Hulsebos T., Schonk D., van Dalen I., Coerwinkel-Driessen M., Schepens J., Ropers H.H. and Wieringa B. (1988) Cytogenet. Cell Genet. 47: 144−148.
15. Jabs E.W. and Persico M.G. (1987) Am. J. Hum. Genet. 41: 374−390.
16. Jabs E.W., Wolf S.W. and Migeon B.R. (1984) Proc. Natl. Acad. Sci. USA 81: 4884−4888.
17. Jabs E.W., Warren A.C., Taylor E.W., Colyer C.R. Meyers D.A. and Antonarakis S.E. (1991) Genomics 9: 141−146.
18. Jones R.S. and Potter S.S. (1985) Nucl. Acids Res. 13: 1027−1042
19. Jorgenson A.L., Bostock C.J. and Bak A.L. (1986) J. Mol. Biol. 187: 185−196
20. Jorgensen A.L., Bostock C.J. and Bak A.L. (1987) Proc. Natl. Acad. Sci. USA 84: 1075−1079.
21. Jorgensen A.L. Kolvraa S., Jones C. and Bak A.L. (1988) Genomics 3: 100−109.
22. Looijenga L.H.J., Smit, V.T.H.B.M., Wessels J.W., Mollevanger P., Oosterhuis, Cornelisse C.J. and Devilee P. (1990) Cytogenet. Cell Genet. 53: 216−218.
23. Manuelidis L. (1978) Chromosoma 66: 23−32
24. McDermid H.E., Duncan A.M.V., Higgins M.J., Hamerton J.L., Rector E., Brasch K.R. and White B.N. (1986) Chromosoma 94: 228−234.
25. Mitchell A.R., Gosden J.R. and Miller D.A. (1985) Chromosoma 92: 369−377.
26. Rocchi M., Baldini A., Archidiacono N., Lainwala, S., Miller O.J. and Miller D.A. (1990) Genomics 8: 705−709.
27. Rocchi M., Archidiacono N., Ward D.C. and Baldini A. (1991) Genomics 9: (in press).
28. Singer M.F. (1982) Int. Rev. Cytol. 76: 67−112.
29. Tyler-Smith C. and Brown W.R.A. (1987) J. Mol. Biol. 195: 457−470
30. Vissel B. and Choo K.H. (1987) Nucl. Acids Res. 15: 6751−6752.
31. Vissel B. and Choo K.H. (1991) Nucl. Acids Res. 19#2: (in press).
32. Warren A.C., Bowcock A.M., Farrer L.A. and Antonarakis S.E. (1990) Genomics 7: 110−114
33. Waye J.S., Creeper L.A. and Willard H.F. (1987a) Chromosoma 95: 182−188.
34. Waye J.S., Durfy S.J., Pinkel D., Kenwrick S., Patterson M., Davies K.E. and Willard H.F. (1987b) Genomics 1:43−51.
35. Waye J.S., England S.B. and Willard H.F. (1987c) Mol. Cell Biol. 7: 349−356
36. Waye J.S., Mitchell A.R. and Willard H.F. (1988) Hum. Genet. 78: 27−32.
37. Waye J.S. and Willard H.F. (1985) Nucl. Acids Res. 13: 2731−2743
38. Waye J.S. and Willard H.F. (1986) Mol. Cell. Biol. 6: 3156−3165.
39. Waye J.S. and Willard H.F. (1987) Nucl. Acids Res. 15: 7549−7569.
40. Waye J.S. and Willard H.F. (1989) Chromosoma 97: 475−480.
41. Willard H.F., Smith K.D. and Sutherland J. (1983) Nucl. Acids Res. 11: 2017−2033.
42. Willard H.F. and Waye J.S. (1987) J. Mol. Evol. 25: 207−214.
43. Wolfe J., Darling S.M., Erickson R.P., Craig I.W., Buckle V.J., Rigby P.W.J., Willard H.F. and Goodfellow P.N. (1985) J. Mol. Biol. 182: 477−485.
44. Yang T.P., Hansen S.K., Oishi K.K., Ryder O.A. and Hamkalo B.A. (1982) Proc. Natl. Acad. Sci. USA 79: 6593−6597.