Brief papers

DNA analysis of Huntington's disease in southern Chinese

Vivian Chan, Y L Yu, T P T Chan, B Yip, C M Chang, M T H Wong, Y W Chan, T K Chan

Abstract

Allelic frequencies of RFLPs at loci closely linked to the HD gene, D4S95, D4S91, D4S141, and D4S90, were determined in 13 Huntington's disease (HD) patients from nine Chinese families and 129 normal subjects. These were similar for non-HD and HD chromosomes and the HD gene in Chinese is associated with multiple haplotypes. Hence the HD gene probably arose independently in the background haplotypes of the Chinese population. The heterozygosity rates for the two most useful RFLP sites are 0·659 for D4S95-*Acc*I VNTR and 0·494 for D4S141-*Hin*dIII.

(CAG)n repeat numbers ranged from 12 to 27 in 174 normal chromosomes. In 52 meiotic recombinations, the (CAG)n repeats were stably inherited in normal families. In HD families, 12 of 13 HD patients had expanded (CAG)n repeats of 40 to 58. Additionally, 10 asymptomatic family members had expanded (CAG)n repeats and the inheritance of the expanded repeat was unstable in these families.

(*J Med Genet* 1995;32:120–124)

University Department of Medicine, Queen Mary Hospital, Hong Kong
V Chan
Y L Yu
T P T Chan
B Yip
T K Chan

Department of Medicine, Ruttonjee Hospital, Hong Kong
C M Chang

University Department of Psychiatry, Queen Mary Hospital, Hong Kong
M T H Wong

Department of Medicine, Kwong Wah Hospital, Kowloon, Hong Kong
Y W Chan

Correspondence to:
Professor V Chan.

Received 16 April 1994
Revised version accepted for publication 12 September 1994

Huntington's disease (HD) is an inherited neurodegenerative disorder manifesting usually in middle life and presenting with involuntary choreic movements, cognitive and psychiatric disturbances, and dementia.[1] There is currently no effective treatment nor means to delay the inexorable progression to death which occurs some 10 to 20 years after the onset of disease. The disease is transmitted in an autosomal dominant manner; although the frequency of new mutation is low,[2] the high degree of penetrance results in large numbers of affected persons within an HD family.

Until recently, various DNA markers from chromosome 4p16.3 have been used in linkage studies of the gene[3–5] and in some cases it was possible to offer predictive testing based on RFLP linkage analysis. The identification of the HD gene (IT 15) and detection at the 5′ end of a trinucleotide repeat sequence (CAG) which is expanded in HD[6] have allowed more accurate molecular diagnosis of the conditiion. Most analyses have been performed on white European and American families.[7 8] The frequency of HD among Japanese is low and linkage analysis supports the view that the gene for Japanese HD may be identical to the western gene.[9] In Hong Kong, where the population is over 95% ethnic southern Chinese, the incidence of HD is low.[10] Previous experience with other genetic diseases, such as β thalassaemia and haemophilia A and B, has shown that the mutations and RFLP heterozygosity rates vary for different ethnic groups.[11–13] We studied allelic frequencies of the various HD DNA markers and analysed trinucleotide repeats in both HD and normal chromosomes in the local Chinese population in an attempt to define the molecular defect in Chinese HD.

Materials and methods

DNA was obtained from nine unrelated HD families consisting of 13 clinically diagnosed HD patients (age of onset ranging from 20 to 70 years) and 36 of their family members. Family members were counselled by a team which included a neurologist, a molecular geneticist, and a psychiatrist. Informed consent was obtained from all the subjects and each was aware of the test being undertaken. DNA samples from 20 β thalassaemia families comprising father, mother, and one child, a haemophilia A family with nine members from three generations, and an additional 30 normal males and 30 normal females (aged 20 to 50 years) were used as controls. All subjects were southern Chinese living in Hong Kong, they or their ancestors being migrants from Guangdong province of South China. Nine unrelated HD chromosomes and 174 unrelated normal chromosomes were available for analysis.

DNA was extracted from peripheral blood leucocytes of each subject. For RFLP analysis, 5 μg amounts were digested with either *Acc*I, *Mbo*I, *Pst*I, *Taq*I, *Msp*I, *Hin*dIII, or *Pvu*II under conditions recommended by the manufacturer (New England Biolabs, Mass, USA). The digested DNA fragments were electrophoresed in 0·8% agarose, transferred onto a nitrocellulose filter, and hybridised with ^{32}P labelled gene probes, using standard techniques.[14] The hybridised filters were washed under stringent conditions and autoradiographed for two to

Table 1 Allele frequencies for RFLPs on non-HD and HD chromosomes

Locus	*Enzyme*	*Allele*	*Non-HD chromosomes*		*HD chromosomes*		*p (Fisher's exact test)*
			No	*%*	*No*	*%*	
D4S95 (674)	*Acc*I	1	160	92	8	89	0·8387
		2	14	8	1	11	
	*Acc*I	1	11	6	1	11	
		2	13	7	1	11	NS†
		3	36	21	3	33	
		4	91	52	3	33	
		5	23	13	1	11	
	*Mbo*I	1	137	79	8	89	0·4080
		2	37	21	1	11	
	*Pst*I	1	153	88	9	100	0·3252
		2	21	12	0	0	
	*Taq*I	1	136	78	4*	50	0·0842
		2	38	22	4	50	
D4S96 (678)	*Msp*I	1	129	74	6	67	0·8157
		2	45	26	3	33	
D4S141 (2R3)	*Hin*dIII	1	79	45	1*	12	0·0664
		2	95	55	7	88	
D4S90 (D5)	*Pvu*II	1	25	14	2	22	0·8693
		2	149	86	7	78	

* One HD patient was excluded from the analysis. He was heterozygous for the site and having no other family member available for study, it was not possible to assign his haplotype.
† $\chi^2 = 1{\cdot}7176$, df = 4, p = 0·7875.

five days. The RPLP sites and probes used were *Acc*I, *Mbo*I, *Pst*I, and *Taq*I detected by probe 674 (1·0 kb *Eco*RI insert) in the D4S95 region,[4] *Msp*I detected by probe 678 (a 0·9 kb *Eco*RI insert) in D4S96,[15] *Hin*dIII detected by probe 2R3 in D4S141,[16] and *Pvu*II hybridised to probe D5 which is a 1·8 kb *Eco*RI insert in the D4S90 region.[17]

AMPLIFICATION OF (CAG)n REPEAT

Genomic primers (HDA and HDB as 5′ and 3′ primers respectively) flanking the trinucleotide repeat sequence were synthesised as described by the Huntington's Disease Collaborative Research Group (1993).[6] This PCR product should include the CAG repeats and the adjacent CCG and CCN repeats between nucleotide (nt) 316-584 of the published sequence.[6] Polymerase chain reaction (PCR) amplification was performed in a 25 µl reaction volume containing 50 ng genomic DNA, 200 mmol/l (each) dNTP, 50 ng of ^{32}P end labelled 5′ primer, 250 ng of unlabelled 5′ primer, 300 ng of unlabelled 3′ primer, 1 U *Taq* Polymerase (Perkin Elmer Cetus), 10% dimethyl sulphoxide, and enzyme buffer as recommended by the manufacturer. After denaturation at 94°C for five minutes, the cycling conditions for 40 cycles were one minute at 94°C, one minute at 60°C, and two minutes extension time at 72°C. A 5 µl aliquot of the PCR product was diluted with an equal volume of 95% formamide loading dye and denatured at 95°C for two minutes before analysis on 8 mol/l urea and 6·5% acrylamide gel (60 cm) at 20 mA. A single stranded M13 was sequenced and analysed in the same gel as standard.

Two other sets of PCR primers were synthesised; HD3 and HD482 according to the sequences published by Kremer *et al.*[18] When HD3 and HD482 were used, the PCR product included the (CAG)n as well as the immediately adjacent set of (CCG)n, between nt 344-482 of the published sequence.[6] PCR cycling conditions were 94°C for one minute, 64°C for one minute, 72°C for two minutes, for 35 cycles. The third 3′ primer used was HD4V, spanning nt 446-431, and when used with HD3 as 5′ primer, the PCR product would only contain the (CAG)n. PCR cycling conditions were 94°C for one minute, 60°C for one minute, and 72°C for two minutes for 35 cycles.

PCR for the trinucleotide repeat associated with dentatorubral and pallidoluysian atrophy (DRPLA) was performed in one HD family, using primers and conditions as described by Nagafuchi *et al.*[19]

Results

RFLP ANALYSIS

The allelic frequencies for the various RFLPs on non-HD (normal) and HD chromosomes are given in table 1. There was no significant difference in allele frequencies between the two groups. Calculation of heterozygosity rate was based on data from the normal chromosomes (table 2). Apart from the highly polymorphic variable number tandem repeats (VNTR) at the *Acc*I site, the *Hin*dIII site in D4S141 would appear the most useful with a heterozygosity rate of 0·494.

On comparing the allelic frequencies of the RFLPs in different ethnic groups (table 3), significant differences were observed in each except for D4S95-*Pst*I, D4S96-*Msp*I, and D4S141-*Hin*dIII sites. With the D4S95-*Taq*I site, the allelic frequency in the Chinese was the reverse of that found in white[4 15 17] and

Table 2 The heterozygosity rate of various RFLPs linked to the HD gene in the Chinese population

Locus (probe)	*Restriction site*	*Frequency*		*Heterozygosity rate (H)*		χ^2
		Allele 1 (p)	*Allele 2 (q)*	*Observed*	*Expected (pq)*	*p*
D4S95 (674)	*Acc*I	0·919	0·080	0·142	0·147	NS†
	*Acc*I (VNTR)	(1) 0·063 (2) 0·074 (3) 0·207	(4) 0·522 (5) 0·132	0·214	0·659*	0·0000‡
	*Mbo*I	0·787	0·212	0·342	0·333	NS
	*Pst*I	0·877	0·122	0·231	0·213	NS
	*Taq*I	0·781	0·218	0·314	0·340	NS
D4S96 (678)	*Msp*I	0·741	0·258	0·157	0·382	0·0028‡
D4S141 (2R3)	*Hin*dIII	0·454	0·545	0·457	0·494	NS
D4S90 (D5)	*Pvu*II	0·143	0·856	0·185	0·244	NS

* Calculated by polynomial equation for multiple alleles, where: $H = 1 - (p^2 + q^2 + r^2 ...)$.
† NS = not significant.
‡ Significant difference between observed and expected heterozygosity rates is because of more people being homozygous for one allele.

Table 3 Allelic frequencies of various RFLPs linked to the HD gene in different ethnic groups

Locus (probe)	Restriction enzyme	Allele	Size (kb)	Frequency (No studied)				χ^2	df	p
				Chinese*	White†	UK‡	Italian§			
D4S95	*Acc*I	1	7·0	0·919 (174)	0·70 (100)	0·58 (99)	–	45·1	2	0·0000
(674)		2	1·5	0·080	0·30	0·42				
	*Acc*I	1	4·5	0·063 (174)	0·07 (100)	–	0·03 (94)	47·4	8	0·0000
		2	4·4	0·074	0·23		0·14			
		3	4·3	0·207	0·10		0·04			
		4	4·2	0·522	0·50		0·78			
		5	4·1	0·132	0·10		0·00			
	*Mbo*I	1	1·2	0·787 (174)	0·65 (100)	0·63 (90)	–	9·4	2	0·0090
		2	0·7/0·5	0·212	0·35	0·37				
	*Pst*I	1	8·6	0·877 (172)	0·91 (100)	0·93 (110)		1·9	2	0·3750
		2	1·7	0·122	0·09	0·07				
	*Taq*I	1	2·3	0·781 (174)	0·38 (100)	0·51 (49)	0·32 (82)	67·7	3	0·0000
		2	1·7	0·218	0·62	0·49	0·68			
D4S96	*Msp*I	1	1·3	0·741 (174)	0·5 (12)	–	–	3·28	1	0·0698
(678)		2	1·0	0·258	0·5					
D4S141	*Hin*dIII	1	5·6	0·454 (174)	–	0·36 (60)	–	1·38	1	0·2388
(2R3)		2	3·9	0·545		0·64				
D4S90	*Pvu*II	1	5·6	0·143 (174)	0·14 (76)	0·90 (96)	–	171·7	2	0·0000
(D5)		2	5·0	0·856	0·86	0·10				

* Chan *et al*: present study, 1994.
† Wasmuth *et al*[4]; Youngman *et al*[17]; Smith *et al*.[15]
‡ Snell *et al*.[5]
§ Frontali *et al*.[20]

Italian[20] populations, while a separate report from the United Kingdom showed approximately equal numbers of both alleles.[5] With the D4S90-*Pvu*II site, the frequencies of alleles 1 and 2 in both Chinese and whites were the reverse of those in the people from the UK.[5]

TRINUCLEOTIDE REPEAT ASSAY

Trinucleotide repeat assays were performed using all three sets of PCR primers. Thus it was possible to calculate the (CAG)n, (CCG)n, and (CCN)n in each sample. The CCG and CCN repeats were not as polymorphic as the CAG repeats, with a range of 5 to 11 repeats for CCG and 6 to 12 repeats for CCN respectively in the 174 normal chromosomes studied and they were not expanded in the HD chromosomes (CCG repeats ranged from 6 to 10, CCN repeats 7 to 13).

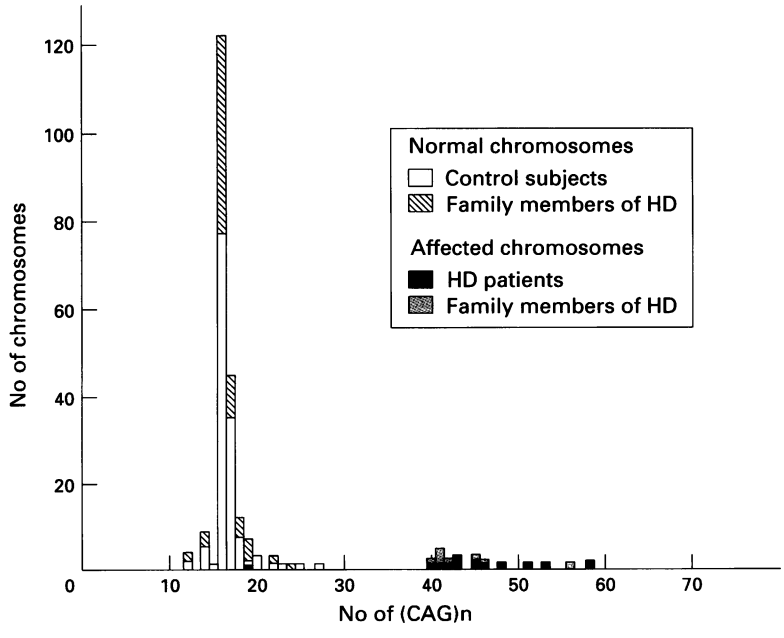

Figure 1 Histogram showing distribution of CAG repeats on the normal chromosomes of control subjects and family members of HD patients, as well as on affected chromosomes of HD patients and asymptomatic family members of these patients with expanded repeats.

(CAG)n REPEAT (USING PRIMERS HD3 AND HD4V)

In 174 normal chromosomes, (CAG)n repeats ranged from 12 to 27 (mean 16·718, SD 1·910). There was no significant difference between males and females (males mean 16·682, SD 2·001 and females mean 16·753, SD 1·830). Values for the affected chromosome in 12 of the 13 clinically diagnosed HD patients ranged from 40 to 58 repeats. In one HD patient, no expanded (CAG)n repeat was observed in four PCR assays. Ten asymptomatic family members were found to have expanded trinucleotide repeats. The distribution of CAG repeats on different chromosomes is shown in fig 1.

In two families, the expanded trinucleotide repeat was inherited by all the sibs, although at the time of study, only the oldest one or two of the sibs had manifested the disease. Fig 2 shows the results of trinucleotide repeat assay in one family. DNA haplotype analysis in this family was also consistent with the conclusion that all sibs carry the abnormal chromosome.

In one family, (CAG)n assay on four different occasions showed that the HD patient had repeats of only 17. These same size repeats were found in one of his two asymptomatic sibs, who, on haplotype analysis, had inherited the same "affected" chromosome as the patient from their dead father, who was affected. The mother who is normal had two alleles of 16 and 17 (CAG)n repeats while the other sister had 16 and 19 (CAG)n repeats. Trinucleotide repeat assay associated with the DRPLA gene showed normal sized repeats for this patient as well as for all his family members.

(CAG)n ANALYSIS IN NORMAL FAMILIES

Determination of the size of trinucleotide repeat in 20 β thalassaemia families as well as the three generations of a haemophilia A family showed that in all 26 offspring (that is, 52

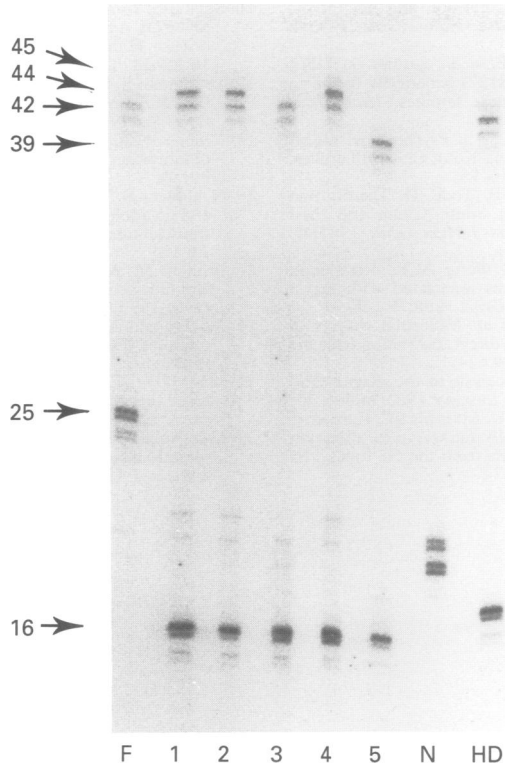

Figure 2 Trinucleotide repeats in an HD family. The father (F) and one of his children (2) are affected. The other asymptomatic children (1, 3, 4, 5) also had expanded trinucleotide repeats. Normal subject (N) and one other patient (HD) were assayed as negative and positive controls. Arrowed numbers indicate repeat sizes of the family.

meiotic recombinations), the (CAG)n repeat lengths were stably inherited from both the paternal and maternal normal genes.

Discussion

The reported incidence of HD in the local Chinese population is low, possibly because within the Chinese culture there is a strong sense of denial of such debilitating disorders, especially when there is no cure. However, with more accurate molecular analysis of the HD gene using trinucleotide repeat assay, predictive testing and prenatal diagnosis could become available and this could have a more positive effect on bringing forth the HD families. In anticipation of the predictive testing service, we studied the (CAG)n repeat of the HD gene in both affected and normal chromosomes and compared this method to RFLP analysis using the various linked DNA markers.

Like previous experiences with β thalassaemia and haemophilia A and B, the heterozygosity rates for the various RFLP sites in Chinese are different from ethnic groups in the West, most notably the *Acc*I site in the D4S95 region, the *Taq*I site in D4S95, and the *Pvu*II site in D4S90. With the latter two RFLP sites, the allele frequencies are the reverse of those found in the other ethinic groups. This underlines the importance of determining the heterozygosity rate for a particular population before deciding on the most suitable site for use in linkage analysis. For southern Chinese, the most useful sites are the D4S95-*Acc*I VNTR and the D4S141-*Hin*dIII site (table 2). In agreement with previous findings of HD in European populations,[21] we have also observed multiple HD haplotypes, even within the small number of patients studied, providing support for the multiple independent origin of mutations causing HD. There was no difference in the allele frequencies of various RFLPs on the normal and the HD chromosomes (table 1).

Trinucleotide repeat of the HD gene in Chinese is similar to that reported for white populations.[7 8 22] There was no overlap in repeat size between the normal group and the affected, possibly because of the relatively small number of clinically diagnosed patients. No potential parental source effect was noted in this group of HD patients and, again, this may be ascribed to the small number studied. The unstable inheritance of the (CAG)n repeat was only observed in HD chromosomes and not in the control group of thalassaemic or haemophiliac families. It is therefore likely that expansion of (CAG)n repeat number in family members is predictive of future disease.

The reason for the absence of the expanded trinucleotide repeat in one HD patient remains unknown. The patient had unfortunately died and as his family refused necropsy, no other sample of blood or brain was available. However, as somatic mosaicism of the expanded HD repeat is thought to be rare,[23] it is unlikely that a brain sample would give a different result from the peripheral blood DNA. CT scanning of this patient showed caudate atrophy and he had psychiatric and cognitive involvement, with typical chorea. His cause of death was pneumonia. DNA haplotyping of this patient and his sister showed they have inherited the same "affected" chromosome, although the sister was still asymptomatic at the time of study. Both the patient and his sister had normal sized (CAG)n repeats associated with the DRPLA gene. It is unlikely that he was suffering from DRPLA. This may represent a different genetic defect for a subgroup of HD patients or that the trinucleotide expansion may only occur in brain tissue. Hence genetic counselling for such a family is not possible at present.

We are grateful to Dr J J Wasmuth, Dr D Shaw, and Professor P S Harper for generous supply of the DNA probes, and the Croucher Foundation for research support.

1 Martin JB, Gusella JF. Huntington's disease: pathogenesis and management. *N Engl J Med* 1986;**315**:1267–76.
2 Harper PS. Genetic aspects of Huntington's disease. In: Harper PS, ed. *Huntington's disease.* Major Problems in Neurology Vol 22. London: Saunders, 1991:281–315.
3 Gusella JF, Wexler NS, Conneally PM, *et al.* A polymorphic DNA marker genetically linked to Huntington's disease. *Nature* 1983;**306**:234–8.
4 Wasmuth JJ, Hewitt J, Smith B, *et al.* A highly polymorphic locus very tightly linked to the Huntington's disease gene. *Nature* 1988;**332**:734–6.
5 Snell RG, Lazarou L, Youngman S, *et al.* Linkage disequilibrium in Huntington's disease: an improved localisation of the genes. *J Med Genet* 1989;**26**:673–5.
6 Huntington's Disease Collaborative Research Group. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* 1993;**72**:971–83.
7 Duyao M, Ambrose C, Myers R, *et al.* Trinucleotide repeat length instability and age of onset in Huntington's disease. *Nature Genet* 1993;**4**:387–92.
8 Snell RG, MacMillan JC, Cheadle JP, *et al.* Relationship

between trinucleotide repeat expansion and phenotypic variation in Huntington's disease. *Nature Genet* 1993;**4**: 393–7.
9 Kanazawa I, Kondo I, Ikeda JE, *et al.* Studies on DNA markers (D4S10 and D4S43/S127) genetically linked to Huntington's disease in Japanese families. *Hum Genet* 1990;**85**:257–60.
10 Wong MTH, Chang PCM, Yu YL, *et al.* Psychosocial impact of Huntington's disease on Hong Kong Chinese families. *Acta Psychiatr Scand* 1994;**90**:16–18.
11 Chan V, Chan TK, Chebab FF, Todd D. Distribution of β-thalassemia mutations in South China and their association with haplotype. *Am J Hum Genet* 1987;**41**: 678–85.
12 Chan V, Chan TK, Liu VWS, Wong ACK. Restriction fragment length polymorphisms associated with factor VIII:C gene in Chinese. *Hum Genet* 1988;**79**:128–31.
13 Chan V, Yip B, Tong TMF, *et al.* Molecular defects in haemophilia B: detection by direct restriction enzyme analysis. *Br J Haematol* 1991;**79**:63–9.
14 Goossens M, Kan YW. DNA analysis in the diagnosis of hemoglobin disorders. *Methods Enzymol* 1981;**76**:805–17.
15 Smith B, Skarecky D, Bengtsson U, Magenis RE, Carpenter N, Wasmuth JJ. Isolation of DNA markers in the direction of the Huntington's disease gene from the G8 locus. *Am J Hum Genet* 1988;**42**:335–44.
16 Snell RG, Youngman S, Lehrach H, Sarfarazi M, Harper PS, Shaw DJ. A new probe (2R3) in the region of Huntington's disease. Human Gene Mapping. *Cytogenet Cell Genet* 1989; **10**:A2621.
17 Youngman S, Shaw DJ, Gusella JF, *et al.* A DNA probe D5 (D4S90) mapping to human chromosome 4p16.3. *Nucleic Acids Res* 1988;**16**:1648.
18 Kremer B, Squitieri F, Telenius H, *et al.* Molecular analysis of late onset Huntington's disease. *J Med Genet* 1993;**30**: 991–5.
19 Nagafuchi S, Yanagisawa H, Sato K, *et al.* Dentatorubral and pallidoluysian atrophy expansion of an unstable CAG trinucleotide on chromosome 12p. *Nature Genet* 1994;**6**: 14–18.
20 Frontali M, Malaspina P, Rossi C, *et al.* Epidemiological and linkage studies on Huntington's disease in Italy. *Hum Genet* 1990;**85**:165–70.
21 MacDonald ME, Novelletto A, Lin C, *et al.* The Huntington's disease candidate gene exhibits many different haplotypes. *Nature Genet* 1992;**1**:99–103.
22 MacMillan JC, Snell RG, Tyler A, *et al.* Molecular analysis and clinical correlations of the Huntington's disease mutation. *Lancet* 1993;**342**:954–8.
23 MacDonald ME, Barnes G, Srinidhi J, *et al.* Gametic but not somatic instability of CAG repeat length in Huntington's disease. *J Med Genet* 1993;**30**:982–6.