

Clustering of Multiallele DNA Markers near the Huntington's Disease Gene

Marcy E. MacDonald,* Shirley V. Cheng,* Michael Zimmer,[‡] Jonathan L. Haines,* Annemarie Poustka,[§] Bernice Allitto,* Barbara Smith,^{||} W. Lance Whaley,* Donna M. Romano,* Jayashree Jagadeesh,* Richard H. Myers,¹ Hans Lehrach,[‡] John J. Wasmuth,^{||} Anna-Maria Frischaut,[‡] and James F. Gusella*

*Molecular Neurogenetics Laboratory, Massachusetts General Hospital and Department of Genetics, Harvard Medical School, Boston, Massachusetts 02114; [‡]Imperial Cancer Research Fund, London, WC2A 3PX, United Kingdom; [§]Zentrum für Molekulare Biologie Heidelberg, D-6900 Heidelberg, Federal Republic of Germany; ^{||}Department of Biological Chemistry, University of California, Irvine, California 92717; ¹Department of Neurology, Boston University Medical School, Boston, Massachusetts 02118

Abstract

Five highly informative multiallele restriction fragment length polymorphisms (RFLPs) of value for preclinical diagnosis of Huntington's disease (HD) have been genetically characterized. One RFLP was uncovered by expansion of the D4S43 locus while three others are at D4S111 and D4S115, loci defined by NotI-linking clones. The final marker, D4S125, represents a recently discovered VNTR locus. All four loci map closer to the HD gene and to the telomere than D4S10, the original linked marker for HD. In combination with two multiallele RFLPs previously identified for D4S43 and another linked locus, D4S95, these five new multiallele markers will dramatically improve the speed and accuracy of predictive testing in HD, and increase its applicability by maximizing the chances of an informative test for anyone with appropriate family structure.

Introduction

Huntington's disease (HD)¹ is a progressive neurodegenerative disorder causing progressive loss of motor control, impaired cognition, and psychiatric disturbance (1). The neuropathologic hallmark of the disorder is marked atrophy of the basal ganglia, particularly the caudate nucleus. HD is caused by an autosomal dominant genetic defect, but the disease gene has not yet been isolated and the biochemical basis for neuronal cell loss in the disorder is not understood. There is no effective treatment for delaying or preventing the manifestations of HD which typically begin in the fourth or fifth decade of life. Linkage analysis with DNA markers has localized the site of the genetic defect to the tip of the chromosome 4 short arm, 3–4 cM distal to the anonymous DNA marker D4S10 (2–4). Two additional polymorphic markers D4S43, and D4S95, have recently been located closer to the telomere than D4S10 and both show very tight linkage with the disease gene (5, 6).

The characterization of linked markers for HD has introduced the possibility of diagnosing the presence of the defect in

gene carriers before the onset of reliable clinical signs (7, 8). The difficult task of making an informed choice concerning predictive testing has prompted the initiation of programs in a number of centers to provide the intensive counselling required. In many cases, however, the modification of genetic risk that can be achieved by DNA typing is limited by family structure, and by the informativeness of the linked markers. For presymptomatic and prenatal diagnosis of HD to be widely applied, it is critically important to increase the number and informativeness of linked DNA markers to ensure a rapid and accurate test. We have therefore searched for new RFLPs at the D4S43 locus (5), and two loci mapped to the HD region by physical techniques, D4S111 and D4S115 (9). In each case, we have identified one or more RFLPs with the characteristics of VNTR (variable number of tandem repeats) markers (10), multiple alleles that can be visualized using any of several restriction enzymes. We have also found that an independently isolated VNTR marker, D4S125 (11), also maps near the HD gene. The availability of a plethora of highly informative multiallele markers should have a profound impact on increasing the accuracy and applicability of presymptomatic and prenatal diagnosis of HD while reducing the time and cost of this stressful procedure.

Methods

DNA probes and hybridization. For the D4S43 locus, two single copy probes were used: (a) pKP1.65 is a 1.7-kb Pst I fragment subcloned from cosmid clone C39 into pGEM3; (b) pXP500 is a single-copy 0.5-kb Xba I-Pst I fragment from cosmid C31 (subcloned into pGEM4) which detects the same multiallele RFLP previously described for probe Sma2.6 (5). p252.3 is a 800 bp Pst I-Sau 3A fragment from a cosmid which overlaps the NotI-linking clone 252 (9) at the D4S115 locus. p157.9 is a 450 bp Pst I-Sau3A fragment subcloned into pGEM3 from a cosmid that overlaps the Not I-linking clone 157 (9) defining D4S111. A second probe for this locus, S281, is a 2.3 kb Eco RI fragment excised from the adjacent cosmid S28. The probes YNZ32 representing D4S125 and pBS674 representing D4S95 have been described previously (6, 11, 12).

Genomic DNA for RFLP typing was extracted from cultured lymphoblasts as previously described (13). Agarose gel electrophoresis, DNA transfer, hybridization, and autoradiography were performed as described (2) except that probes were labeled by the oligonucleotide priming method of Feinberg and Vogelstein (14).

Genetic analysis. RFLPs were identified by hybridization of the probes to Southern blots containing DNAs from five unrelated individuals digested with each of 15–30 restriction enzymes. Individual variations in the pattern of bands observed were tested for Mendelian inheritance by typing large sibships of the Venezuela reference pedigree (15). The multiallele RFLPs showed similar variation with many different restriction enzymes. Heterozygosity in the normal population was determined by typing 39 randomly chosen North American cau-

Address reprint requests to Dr. J. F. Gusella, Director, Molecular Neurogenetics Laboratory, Massachusetts General Hospital, Boston, MA 02114.

Received for publication 20 March 1989.

1. *Abbreviations used in this paper:* HD, Huntington's disease; RFLP, restriction fragment length polymorphism; VNTR, variable number of tandem repeats.

J. Clin. Invest.

© The American Society for Clinical Investigation, Inc.

0021-9738/89/09/1013/04 \$2.00

Volume 84, September 1989, 1013–1016

casians and 61 unaffected parents of HD patients, though not all individuals were typed for all marker loci. Heterozygosity in the HD population was determined by typing the 61 unrelated HD patients. The estimate of linkage distance between each marker locus and the HD gene was obtained by typing the Venezuela HD kindred and several smaller pedigrees from the United States and Europe (2, 3, 16). Lod scores indicating the log of the relative likelihood of linkage were calculated using the LINKAGE program (17). An age of onset correction was used for all calculations involving the HD locus (2, 18). The maximum lod score (\hat{z}) denotes the maximum likelihood estimate $\hat{\theta}$ of the recombination frequency between the marker and the disease gene. The 1-lod unit confidence interval on this estimate of $\hat{\theta}$ approximates the 95% confidence limits (19). To maximize the linkage information for the D4S43 locus, previously reported RFLPs were included in a haplotype with the multiallele RFLPs, including a very rare *Sau* 96I RFLP segregating with HD in the Venezuela kindred (5). Since D4S111 and D4S115 are 250 kb apart (9) and are extremely tightly linked, with no recombinants in 179 informative meioses in reference pedigrees, the two were haplotyped as a single locus for this analysis.

Results

The mapping of DNA probes using specially constructed somatic cell hybrid panels (12, 20) has provided a number of loci distal to D4S10 in 4p16.3, the terminal cytogenetic subband of the chromosome 4 short arm. Two of these, D4S43 and D4S95 have been shown to be more tightly linked to HD than D4S10 (5, 6). In this study, we have concentrated on expanding the previously described D4S43 locus, and on finding RFLPs for D4S111 and D4S115, two loci representing *Not*I-linking clones mapped by physical means distal to D4S10 and D4S43 (9). We have also used the somatic cell hybrid mapping panel to assign D4S125 (11), an independently isolated VNTR marker, to the terminal 4p16.3 band.

The D4S43 locus was originally defined by the single copy probe C4H chosen from a λ phage library enriched for chromosome 4 sequences (21). The region was then expanded to ~ 108 kb by cosmid walking, permitting the identification of a

number of RFLPs of limited informativeness, primarily 2-allele systems (5). One RFLP, detected by the slightly repetitive probe *Sma*2.6 (and more easily by the recently derived single-copy subclone pXP500), provided most of the polymorphism information content (22) of the locus because it had the characteristics of a VNTR RFLP with four distinguishable alleles. We have continued cosmid walking in the same direction from C4H, and the size of the cloned region has now reached 180 kb. Within the additional cosmids, we screened for RFLPs by hybridization of single-copy probes to DNA from five unrelated individuals digested with 15–30 different restriction enzymes. One probe, pKP1.65, derived from a site 35 kb beyond *Sma*2.6, also detected an insertion/deletion pattern with numerous enzymes (Table I).

Similar but independent VNTR-like RFLPs were found when probes from cosmids overlapping two *Not*I-linking clones were used (Table I and Fig. 1). These probes p157.9 and p252.3 define the loci D4S111 and D4S115, respectively, which are located 250 kb apart, and are both distal to D4S43. A second multi-allele RFLP for D4S111 was detected by the probe pS281, derived from an adjacent cosmid clone. Since D4S95 also reveals a multiallele RFLP, this study brings to seven the number of such polymorphisms located in 4p16.3. The high frequency of VNTR markers identified without any means of enrichment suggests that such polymorphisms may be preferentially located toward telomeres.

In any given predictive test, it is crucial that the affected parent of the "at risk" individual be heterozygous for the marker locus. To estimate the frequency with which individuals carrying the HD gene are heterozygous for one or more of the multiallele systems, we have typed 61 unrelated HD patients. All 61 individuals were heterozygous for at least one marker. To assess heterozygosity in the normal population, we typed the unaffected parents of the HD individuals, along with 39 randomly chosen North Americans from non-HD families. None of these 100 non-HD individuals was homozygous for all seven multiallele RFLPs indicating that the capacity to distinguish any two chromosome 4s approaches 100%. Like previously characterized RFLPs in the 4p16.3 region, none of

Table I. Description of Four New Multiallele Polymorphisms in 4p16.3

Probe name	Locus symbol	Enzyme*	Allelic fragments†
pKP1.65	D4S43	<i>Stu</i> I	>6 alleles from 2.0–2.3 kb; 110 chromosomes screened; heterozygosity = 0.53.
p157.9	D4S111	<i>Pst</i> I	>4 alleles from 1.8–2.1 kb; 116 chromosomes screened; heterozygosity = 0.69.
pS281	D4S111	<i>Hinc</i> II	>4 alleles from 6.6–7.2 kb; 56 chromosomes screened; heterozygosity = 0.22.
p252.3	D4S115	<i>Pst</i> I	>6 alleles from 2.2–2.6 kb; 120 chromosomes screened; heterozygosity = 0.60.

* Each RFLP is detectable with many restriction enzymes. The enzyme shown has been used most extensively.

† None of the probes detects invariant fragments for the indicated restriction enzyme.

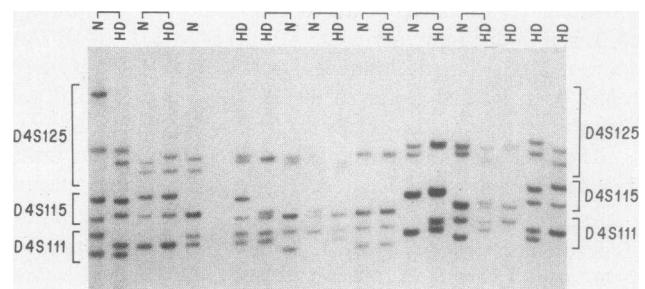


Figure 1. Simultaneous hybridization with three probes detecting multiallele RFLPs near the Huntington's disease gene. The potential for simultaneous detection of RFLPs at the loci D4S111, D4S115 and D4S125 was tested by hybridization of probes for these loci to *Pst*I-digested DNAs from 11 individuals with Huntington's disease (HD) together with 8 normal parents (N). All individuals were heterozygous for at least one RFLP. The HD gene was not consistently associated with any single allele at any of the three marker loci. For pairs of HD and N lanes, the allele at each locus transmitted with HD can usually be inferred by discounting the shared allele.

Table II. Pairwise Lod Scores for Linkage of 4p16.3 Markers to HD

Locus 1	Locus 2	Recombination fraction (θ)							Maximum lod score (\hat{z})	Confidence interval (1-lod unit)
		0.00	0.01	0.05	0.10	0.20	0.30	0.40		
HD	D4S43	$-\infty$	47.1	45.2	41.4	31.9	21.0	9.6	47.1 $\hat{\theta} = 0.01$	$\theta = .001-0.04$
HD	D4S125	$-\infty$	24.4	23.7	21.4	15.6	9.2	3.2	24.5 $\hat{\theta} = 0.02$	$\theta = .001-0.05$
HD	D4S111/S115	$-\infty$	21.0	21.6	19.8	14.5	8.6	3.0	21.8 $\hat{\theta} = 0.03$	$\theta = .005-0.08$

the new RFLPs showed any allele preference on HD chromosomes that would indicate significant linkage disequilibrium with the defect.

For the markers to be effectively used in preclinical HD diagnosis, we have estimated the genetic distance of each locus from the disease gene by typing the polymorphisms in several HD pedigrees, including a very large Venezuela HD kindred (2). The results of this analysis are given as pairwise lod scores in Table II. This analysis indicates that each of the newly described loci is tightly linked to the disease gene, with predicted recombination estimates ranging from 1 to 3 cM.

A major benefit of multiallele systems is the potential for typing several loci using the same restriction enzyme. Given the stress on the at risk individual and his family imposed by the presymptomatic testing process, the turn-around time for DNA typing should ideally be as short as possible. Consequently, we have tested the feasibility of simultaneous hybridization of several probes to the same blots for more rapid processing of test samples. Fig. 1 shows the results of hybridization of a probe mixture containing pYNZ32 [D4S125], p252.3 [D4S115], and p157.9 [D4S111] to Pst I-digested DNA from a panel of HD victims and their normal relatives. The three RFLPs were easily distinguished in this simultaneous analysis, with every individual heterozygous for at least one of the markers. This combination of loci provides a highly informative rapid initial screen for heterozygosity. In those rare instances where none of these markers is informative for pre-clinical testing, various combinations of the remaining four multiallele markers can be applied.

Discussion

The linked marker D4S10 has been applied successfully to presymptomatic diagnosis of HD, but the accuracy of prediction has been limited by the measurable recombination between the marker and HD, and by the lack of a flanking marker (8). A further complication of utilizing two allele systems is their limited capacity to detect nonpaternity in the ragged incomplete pedigrees typical of at risk individuals requesting preclinical testing. Finally, the procedure using D4S10 is costly and tedious, since it requires typing family members for many different RFLP sites with several different restriction enzymes to maximize the potential for an informative test. The addition of RFLPs for D4S43 and D4S95 has increased the applicability of the predictive test, but even so, many cases still yield equivocal or uninformative results, further complicating the already difficult task of genetic counseling in this situation.

The characterization of the five multiallele RFLPs described here promises to alleviate these difficulties. The

markers can be typed using any of a number of restriction enzymes, some of which are in common for each other and for certain D4S10 RFLPs. Thus, sequential or even simultaneous hybridization of a single Southern blot can potentially replace the numerous blots required previously. The high level of heterozygosity observed with these markers will also increase the proportion of tests that are informative and accelerate the testing process. Moreover, the fact that all the new RFLPs are distal to D4S10 and much closer to HD confers a significant increase in the predictive accuracy of test results.

The relative order of these loci on the physical map of 4p16.3 is already known from pulsed-field gel mapping experiments, although the precise distances separating the markers has not been determined in all cases (Bucan et al., manuscript submitted for publication). Similarly, some of the loci have been genetically ordered by linkage analysis in reference pedigrees (23). The combined information from these two complementary methods indicates the following linear order: centromere-D4S10-D4S125-D4S95-D4S43-D4S115-D4S111-telomere with the region D4S125 to D4S111 spanning only ~ 2 cM.

Typing of these markers against individual recombination events in HD pedigrees indicates that the disease gene is distal to D4S125, D4S95, and D4S43 (24) but cannot be unequivocally placed relative to D4S115 and D4S111 (23). Two recombination events suggest a terminal location for HD while one supports the view that D4S111 and D4S115 might flank the disease gene on the telomere side. Until this ambiguity has been resolved, these latter loci should not be assumed to flank the disease gene. We urge that at present a conservative recombination estimate of 3% be used for all of these markers when they are used in genetic counseling.

Acknowledgments

We thank the many clinicians who supplied HD pedigrees, particularly members of the Venezuela Collaborative Huntington's Disease Project. We are grateful to Dr. Y. Nakamura and Dr. R. White (Howard Hughes Medical Institute) for supplying the pYNZ32 probe and to Dr. Charles Cantor, Dr. Francis Collins, Dr. David Housman, Dr. Cassandra Smith, Dr. Allan Tobin, and Dr. Nancy Wexler, our colleagues in the Hereditary Disease Foundation Collaborative HD Research Group, for their helpful discussions.

This work was supported by National Institute for Neurological Communicative Disorders and Stroke grants NS16367 (Huntington's Disease Center Without Walls), NS22031, and NS20012, and by grants from the Hereditary Disease Foundation and the Huntington's Disease Society of America. Drs. Cheng, Whaley, and Allitto received postdoctoral fellowships from the Hereditary Disease Foundation, the Huntington's Disease Society of America, and the Wills Foundation, respectively.

References

1. Martin, J. B., and J. F. Gusella. 1986. Huntington's disease: Pathogenesis and management. *N. Engl. J. Med.* 315:1267-1276.
2. Gusella, J. F., N. S. Wexler, P. M. Conneally, S. L. Naylor, M. A. Anderson, R. E. Tanzi, P. C. Watkins, K. Ottina, M. R. Wallace, A. Y. Sakaguchi, A. B. Young, I. Shoulson, E. Bonilla, and J. B. Martin. 1983. A polymorphic DNA marker genetically linked to Huntington's disease. *Nature (Lond.)*. 306:234-238.
3. Gusella, J. F., R. E. Tanzi, M. A. Anderson, W. Hobbs, K. Gibbons, R. Raschtchian, T. C. Gilliam, M. R. Wallace, N. S. Wexler, and P. M. Conneally. 1984. DNA markers for nervous system disorders. *Science (Wash. DC)*. 225:1320-1326.
4. Gilliam, T. C., R. E. Tanzi, J. L. Haines, T. I. Bonner, A. G. Faryniarz, W. J. Hobbs, M. E. MacDonald, S. V. Cheng, S. E. Folstein, P. M. Conneally, N. S. Wexler, and J. F. Gusella. 1987. Localization of the Huntington's disease gene to a small segment of chromosome 4 flanked by D4S10 and the telomere. *Cell*. 50:565-571.
5. Gilliam, T. C., M. Bucan, M. E. MacDonald, M. Zimmer, J. L. Haines, S. V. Cheng, T. M. Pohl, W. L. Whaley, B. A. Allitto, A. Faryniarz, J. J. Wasmuth, A. M. Frischauf, P. M. Conneally, H. Lehrach, and J. F. Gusella. 1987. A DNA segment encoding two genes very tightly linked to Huntington's disease. *Science (Wash. DC)*. 238:950-952.
6. Wasmuth, J. J., J. Hewitt, B. Smith, D. Allard, J. L. Haines, D. Skarecky, E. Partlow, and M. R. Hayden. 1988. A highly polymorphic locus very tightly linked to the Huntington's disease. *Nature (Lond.)*. 332:734-736.
7. Wexler, N. S., P. M. Conneally, and J. F. Gusella. 1984. A DNA polymorphism for Huntington's disease marks the future. *Arch. Neurol.* 42:20-24.
8. Meissen, G. J., R. H. Myers, C. A. Mastromauro, W. J. Koroshetz, K. W. Klinger, L. A. Farrer, P. A. Watkins, J. F. Gusella, E. D. Bird, and J. B. Martin. 1988. Predictive testing for Huntington's disease with use of a linked DNA marker. *N. Engl. J. Med.* 318:535-542.
9. Pohl, T. M., M. Zimmer, M. E. MacDonald, B. Smith, M. Bucan, A. Poustka, S. Volinia, S. Searle, G. Zehetner, J. J. Wasmuth, J. Gusella, H. Lehrach, and A. M. Frischauf. 1988. Construction of a NotI linking library and isolation of new markers close to the Huntington's disease gene. *Nucleic Acids Res.* 16:9185-9198.
10. Nakamura, Y., M. Leppert, P. O'Connell, R. Wolff, T. Holm, M. Culver, C. Martin, E. Fujimoto, M. Hoff, E. Kumlin and R. White. 1987. Variable number of tandem repeat (VNTR) markers for human gene mapping. *Science (Wash. DC)*. 235:1616-1622.
11. Nakamura Y., M. Culver, P. O'Connell, M. Leppert, G. M. Lathrop, J. M. Lalouel, and R. White. 1988. Isolation and mapping of a polymorphic DNA sequence (pYNZ32) on chromosome 4p [D4S125]. *Nucleic Acids Res.* 16:4186.
12. Smith, B., D. Skarecky, U. Bengtsson, R. E. Magenis, N. Carpenter, and J. J. Wasmuth. 1988. Isolation of DNA markers in the direction of the Huntington disease gene from the G8 locus. *Am. J. Hum. Genet.* 42:335-344.
13. Gusella, J. F., A. Varsanyi-Breiner, F. T. Kao, C. Jones, T. T. Puck, C. Keys, S. Orkin, and D. E. Housman. 1979. Precise localization of the human β globin gene complex on chromosome 11. *Proc. Natl. Acad. Sci. USA*. 76:5239-5243.
14. Feinberg, A. P. and B. Vogelstein. 1984. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity: Addendum. *Anal. Biochem.* 137:266-267.
15. Tanzi, R. E., J. L. Haines, P. C. Watkins, G. D. Stewart, M. R. Wallace, R. Hallowell, C. Wong, N. S. Wexler, P. M. Conneally, and J. F. Gusella. 1988. Genetic linkage map of human chromosome 21. *Genomics*. 3:129-136.
16. Haines, J., R. Tanzi, N. Wexler, P. Harper, S. Folstein, J. Cassiman, R. Myers, A. Young, M. Hayden, A. Falek, E. Tolosa, S. Crespi, L. DiMaio, G. Holmgren, M. Anvret, I. Kanazawa, J. Gusella, and M. Conneally. 1986. No evidence of linkage heterogeneity between Huntington disease (HD) and G8 (D4S10). *Am. J. Hum. Genet.* 39:A156.
17. Lathrop, G. M., J. M. Lalouel, C. Julier, and J. Ott. 1984. Strategies for multilocus linkage analysis in humans. *Proc. Natl. Acad. Sci. USA*. 81:3443-3446.
18. Hodge, S. E., L. A. Morton, S. Tidemam, K. K. Kidd, and M. A. Spence. 1979. Age of onset correction available for linkage (LIPED). *Am. J. Hum. Genet.* 31:761-762.
19. Conneally P. M., J. H. Edwards, K. K. Kidd, J. M. Lalouel, N. E. Morton, J. Ott, and R. White. 1985. Report of the Committee on Methods of Linkage Analysis and Reporting. *Cytogenet. Cell Genet.* 40:356-359.
20. MacDonald, M. E., M. A. Anderson, T. C. Gilliam, L. Tranebjærg, N. J. Carpenter, E. Magenis, M. R. Hayden, S. T. Healey, T. I. Bonner, and J. F. Gusella. 1987. A somatic cell hybrid panel for localizing DNA segments near the Huntington's disease gene. *Genomics*. 1:29-34.
21. Gilliam, T. C., S. T. Healey, M. E. MacDonald, J. Wasmuth, G. D. Stewart, J. C. Roy, and J. F. Gusella. 1987. Isolation of polymorphic DNA from human chromosome 4. *Nucleic Acids Res.* 15:1445-1458.
22. Botstein, D., R. L. White, M. Skolnick, and R. Davis. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* 32:314-331.
23. MacDonald, M. E., J. L. Haines, M. Zimmer, S. V. Cheng, S. Youngman, W. L. Whaley, M. Bucan, B. A. Allitto, N. Wexler, B. Smith, et al. Recombination events suggest potential sites for the Huntington's disease gene. *Neuron*. In press.
24. Whaley, W. L., F. Michiels, M. E. MacDonald, D. Romano, M. Zimmer, B. Smith, J. Leavitt, M. Bucan, J. Haines, T. C. Gilliam, G. Zehetner, A. M. Frischauf, J. J. Wasmuth, H. Lehrach, and J. F. Gusella. 1988. Mapping of D4S98/S114/S113 confines the Huntington's defect to a reduced physical region at the telomere of chromosome 4. *Nucleic Acids Res.* 16:11769-11780.