Proc. Natl. Acad. Sci. USA
Vol. 86, pp. 10011–10014, December 1989
Genetics

The Huntington disease locus is most likely within 325 kilobases of the chromosome 4p telomere

(physical mapping/reverse genetics/pulsed-field gel electrophoresis)

NORMAN A. DOGGETT*†, JAN-FANG CHENG*‡, CASSANDRA L. SMITH§‡, AND CHARLES R. CANTOR*‡

*Department of Genetics and Development and §Departments of Microbiology and Psychiatry, College of Physicians and Surgeons, Columbia University, 701 West 168th Street, New York, NY 10032

Contributed by C. R. Cantor, September 12, 1989

ABSTRACT **The genetic defect responsible for Huntington disease was originally localized near the tip of the short arm of chromosome 4 by genetic linkage to the locus D4S10. Several markers closer to Huntington disease have since been isolated, but these all appear to be proximal to the defect. A physical map that extends from the most distal of these loci, D4S90, to the telomere of chromosome 4 was constructed. This map identifies at least two CpG islands as markers for Huntington disease candidate genes and places the most likely location of the Huntington disease defect remarkably close (within 325 kilobases) to the telomere.**

Huntington disease (HD), an autosomal-dominant neurodegenerative disease of the basal ganglia, is characterized by progressive loss of motor control and a gradual impairment of intellectual and emotional capabilities (1–3). The disease typically starts in midlife and results in death within 15–20 years. The biochemical defect responsible for the disease is unknown. Attempts to isolate the HD gene have relied on determining its chromosomal location.

The HD gene was first localized distal to the polymorphic marker G8 (locus D4S10) in the last cytogenetic subband of chromosome 4 (4p16.3) (refs. 4–6). Extensive cloning efforts in this region uncovered several polymorphic markers that have placed HD progressively closer to the telomere (7–13). Recent evidence from family studies suggests that HD may, in fact, be telomeric to the most distal locus, D4S90 (clone D5) (14). We have constructed a physical map extending from the telomere of 4p through D4S90. This map severely limits the region most likely to contain the HD gene and shifts the focus of the HD gene search to attempts to locate specific candidate genes.

MATERIALS AND METHODS

Cell Lines. HHW693 is a hamster–human hybrid cell line, retaining only the p arm of a 4/5 translocation chromosome (5cen-5p15.1::4p15.1-4pter) (15). HHW416 is a hamster–human hybrid cell line containing an intact human chromosome 4 (16). UCW56 is the parental cell line for both hybrids.

Preparation of DNA Inserts and Restriction Enzyme Digests. Inserts (100 μl) were prepared from 10^7 cells per ml of low-melting agarose as described (17, 18). Complete restriction digests were performed with 20 units of enzyme per μg of DNA for 4 hr at 37°C. One-sixth slice of each insert ($\approx$1 μg of DNA) was loaded in each pulsed-field gel (PFG) lane. Phage λ DNA ladders were prepared as reported (19) and used as size markers for PFG electrophoresis.

PFG Electrophoresis. PFG electrophoresis was performed on a Pulsaphor apparatus (Pharmacia LKB) using the double inhomogeneous electrode configuration as described (20), except for Fig. 2*A* where the hexagonal electrode kit was used. Gels were cast from 1.2% agarose (SeaKem LE Agarose; FMC) and were electrophoresed at 15°C in modified TBE buffer (100 mM Tris base/100 mM boric acid/0.2 mM EDTA, pH 8.2). Pulse times, electrophoresis times, and the voltage employed for each experiment are indicated.

Southern Blotting. The DNA in agarose gels was UV-irradiated at 310 nm (LKB Macrovue Transilluminator) for 90–120 s to introduce nicks, then denatured in 0.5 M NaOH/0.5 M NaCl for 1 hr, and neutralized in 0.5 M Tris·HCl, pH 7.5/0.5 M NaCl for 1 hr. DNA was transferred to nylon membranes (Nytran; Schleicher & Schuell) by capillary action with 15× SSC (1× = 0.15 M NaCl/0.015 M sodium citrate, pH 7.0).

Probe Preparation and Hybridization. Plasmid inserts isolated from agarose gels were ^{32}P-labeled by priming with random oligomers (21). In Fig. 1, prehybridization, hybridization, and washing were performed as described (22). For all other experiments, prehybridization was for 2–4 hr at 42°C in 50% (vol/vol) formamide/5× SSC/5× Denhardt's solution/50 mM sodium phosphate, pH 6.8/0.1% SDS/denatured salmon sperm DNA (100 μg/ml). (1× Denhardt's solution = 0.02% polyvinylpyrrolidone/0.02% Ficoll/0.02% bovine serum albumin.) Hybridization was for 14–18 hr in 50% formamide/6× SSC/1× Denhardt's solution/20 mM sodium phosphate, pH 6.8/0.1% SDS/10% (wt/vol) dextran sulfate/heparin (500 μg/ml)/denatured salmon sperm DNA (100 μg/ml). Filters were washed twice in 2× SSC at room temperature and then four times in either 2× SSC at 68°C (for Pvu0.6) or 0.5× SSC at 68°C (for D4S90). The filters were blotted semi-dry and exposed to x-ray film (Kodak XAR-5) at −70°C with two intensifying screens (DuPont Cronex Lightning Plus) for 3–10 days.

RESULTS AND DISCUSSION

We have described (22) the cloning of a human telomeric *Eco*RI fragment in a half yeast artificial chromosome vector. This clone, YHT1, contains a 600-base-pair *Pvu* I fragment (Pvu0.6) that is common to human but not rodent telomeres. We have used this fragment to identify and characterize the p arm telomere of chromosome 4. High molecular weight DNA from HHW693, a hybrid cell line containing the 4p telomere (15), was digested with BAL-31 exonuclease, then treated with the restriction enzyme *Bgl* II, and subjected to Southern blot analysis. In the absence of BAL-31 treatment, probe Pvu0.6 detected a single broad genomic DNA fragment (Fig. 1*A*, lane 1). Its fuzzy appearance is characteristic of telomeric restriction fragments and presumably reflects the

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: HD, Huntington disease; PFG, pulsed-field gel.

†Present address: Los Alamos National Laboratory, LS-3, MS-M886, Los Alamos, NM 87567.
‡Present address: Human Genome Center, Lawrence Berkeley Laboratory, University of California, Berkeley, CA 94720.

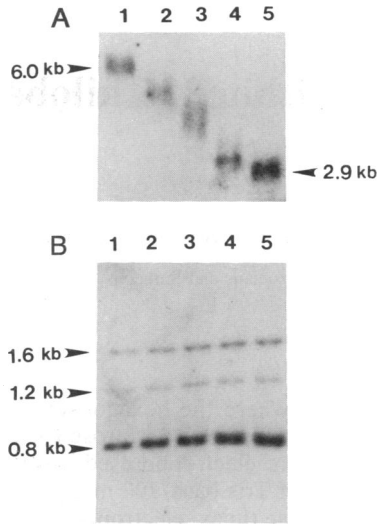

FIG. 1. Effect of BAL-31 digestion on *Bgl* I fragments from cell line HHW693 hybridizing to the telomeric probe Pvu0.6 (*A*) and to probe D5 (*B*). Genomic DNA (5 μg) was treated with 1 unit of BAL-31 nuclease at 30°C for 20, 40, 60, and 120 min as described (15) (lanes 2, 3, 4, and 5, respectively). Lane 1 contains an untreated control. All samples were subsequently digested with *Bgl* I, electrophoresed in a 1.0% agarose gel, and transferred to a nylon membrane. This blot was first hybridized to probe Pvu0.6 (*A*), stripped, and rehybridized with probe D5 (*B*).

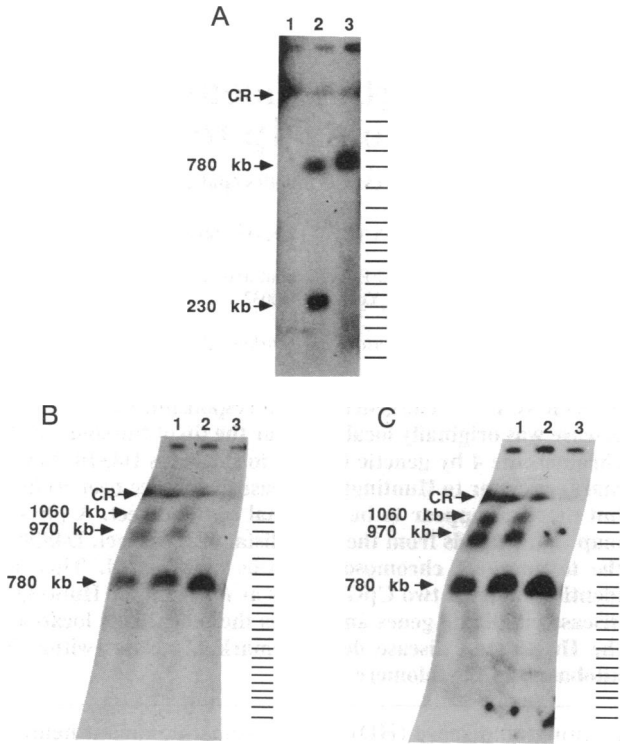

FIG. 2. Identification of the telomeric *Not* I fragments on chromosome 4 by PFG electrophoresis. (*A*) Hybridization of telomeric probe Pvu0.6 to complete *Not* I digests of hamster–human hybrid cell line DNA fractionated by PFG electrophoresis. Lanes: 1, UCW56 hamster parental cell line; 2, HHW416 hybrid containing intact chromosome 4; 3, HHW693 hybrid containing the tip of chromosome 4p. (*B*) PFG electrophoresis-fractionated *Not* I partial digest of HHW693 genomic DNA hybridized to probe Pvu0.6. Lanes: 1, 2, and 3, digested with 0.05, 0.1, and 0.5 unit/μg of DNA for 2.5 hr at 37°C, respectively. (*C*) Same blot shown in *B* rehybridized with probe D5. CR, compression region containing large DNA that is not resolved. The positions of annealed phage λcI_{857} DNA concatemer size standards (monomer = 48.5 kb) are indicated by solid lines. PFG electrophoresis was carried out for 48 hr at 160 V with 125-s pulse times in *A* and for 44 hr at 300 V with 150-s pulse times in *B* and *C*.

dynamic nature of the growing telomere. With increasing BAL-31 digestion times the 6-kilobase (kb) band progressively shortened to approximately 3 kb (lanes 2–5), confirming its telomeric location. When this same blot was rehybridized with D5, three bands, all insensitive to BAL-31 were detected (Fig. 1*B*). Since there are no *Bgl* II sites in probe D5, the detection of multiple *Bgl* II bands within a single *Not* I fragment of cell line HHW693 (see below) suggests that D5 contains a member of a cluster of low copy number repeats present on chromosome 4p.

The size of the terminal *Not* I fragment of chromosome 4 was determined in cell lines HHW693, HHW416 (a hybrid cell line containing an intact chromosome 4, ref. 16), and UCW56 (the hamster parental line). *Not* I-digested DNA from these cell lines was fractionated by PFG electrophoresis (23), and hybridized to probe Pvu0.6 (Fig. 2*A*). No hybridization signal was detected in DNA from the parental hamster line UCW56 (lane 1). A single 780-kb band from the end of the 4p arm was detected for HHW693 (lane 3). This fragment was also detected in cell line HHW416 as well as an additional 230-kb *Not* I fragment, presumably the 4q telomere (lane 2). The 780-kb telomeric *Not* I fragment of 4p was the same size as the *Not* I fragment detected with probe D5 (data not shown).

Are the two 780-kb fragments the same or just the coincidental migration of two similar size fragments? To answer this question probes Pvu0.6 and D5 were hybridized separately to the same *Not* I partial digest of HHW693 DNA (Fig. 2 *B* and *C*). Besides the 780-kb band, the same 970-kb and 1060-kb bands were detected by each probe, confirming that locus D4S90 is located on the telomeric *Not* I fragment.

Several additional restriction enzyme digests of HHW693 DNA were analyzed by PFG electrophoresis to limit, further, the distance between locus D4S90 and the telomere and to develop a more detailed map of this region. Filters were hybridized to probe Pvu0.6 (Fig. 3 *A* and *B*), stripped, and then rehybridized with D5 (Fig. 3 *C* and *D*). The physical map derived from these hybridization results is shown in Fig. 4. Probe Pvu0.6 hybridized to a single 325-kb *Nru* I fragment, the predominant fragment also detected with D5 (Fig. 3 *A* and *C*). Thus D5 and presumably HD are within 325 kb of the telomere. Probe D5 detected an additional strong 150-kb *Nru* I band plus a very weak 105-kb band (which probably arose from cross-hybridization or from partial cleavage within the 150-kb band). The 150-kb *Nru* I fragment is adjacent to the terminal 325-kb *Nru* I fragment, because PFG electrophoresis analysis of an *Nru* I partial digest (data not shown) reveals two *Nru* I fragments, 325 kb and 475 kb, that hybridized to probe Pvu0.6. Probe Pvu0.6 detects a 230-kb *Sal* I fragment that hybridized weakly to probe D5. The simplest interpretation is that virtually all of the DNA recognized by probe D5 lies between the *Sal* I and distal *Nru* I sites, 230–325 kb from the telomere. In a separate *Sal* I partial digest, the major *Sal* I fragment, 450 kb, detected by D5 was shown to be adjacent to the telomeric *Sal* I fragment (data not shown).

Other restriction fragments detected with D5 could not be positioned relative to the telomere because no clear size overlap was seen with fragments detected by Pvu0.6. However, considerable insights about the structure of the telomere were provided using probe Pvu0.6. Restriction enzymes that cleave at A+T-rich sites, such as *Ase* I, *Dra* I, and *Ssp* I, produced weak diffuse telomeric bands much smaller than 10 kb that are not readily visible in these PFG separations (Fig. 3 *A* and *B*). This is consistent with the A+T-rich content of DNA near telomeres (22). In contrast, most restriction enzymes recognizing CpG sequences yielded larger frag-

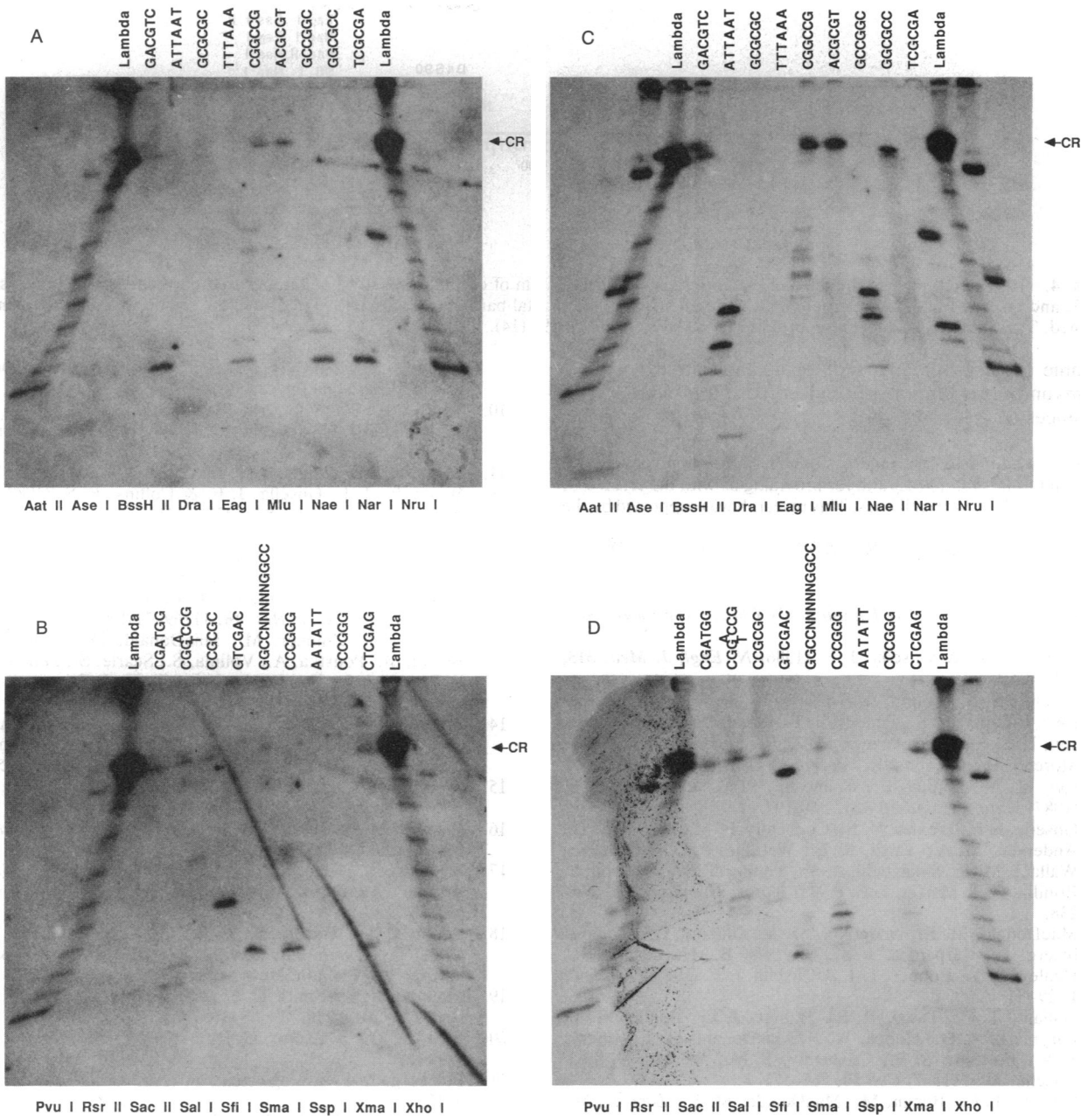

FIG. 3. Restriction enzyme digests of the telomere region of the p arm of chromosome 4. (*A* and *B*) Hybridization of telomeric probe Pvu0.6 to PFG electrophoresis-fractionated HHW693 DNA digested with several restriction enzymes. (*C* and *D*) The blots shown in *A* and *B* were stripped and rehybridized with probe D5. The restriction enzymes used are indicated below each lane, and their corresponding recognition sequences are shown above each lane. Annealed phage λcI_{857} (48.5 kb) concatemers were included in the outside lanes as size standards. CR, compression region containing DNA greater than 550 kb. PFG electrophoresis was carried out for 40 hr at 300 V at 14°C with 50-s pulse times.

ments that were defined by two unmethylated CpG-rich regions (CpG islands) (24, 25). The restriction enzymes *Bss*HII, *Eag* I, *Nae* I, *Nar* I, *Rsr* II, and *Sac* II all produced telomeric DNA fragments close to 50 kb (Figs. 3 *A* and *B* and 4). Additionally, *Eag* I, *Nae* I, *Sac* II, *Sfi* I, *Sma* I, and *Xma* I sites were clustered 100 kb from the telomere, indicating that this region contained a second CpG island (Figs. 3 *A* and *B* and 4). These telomeric restriction sites were usually detected in nominally complete digests; however, in some cases they were revealed by the presence of multiple bands, presumably the result of partial methylation. More extensive partial digestions probed with Pvu0.6 should eventually allow the positioning of any additional CpG islands on the telomeric *Not* I fragment.

Available evidence suggests that the gene responsible for HD is distal to locus D4S90 (clone D5) (14). The results presented here indicate that HD is within 325 kb of the telomere of chromosome 4p. Furthermore, they reveal the presence of two dense clusters of restriction enzyme cleavage sites containing unmethylated CpG sequences in the distal 100 kb of the chromosome. Such sequences have been found to be preferentially located near the 5′ regions of genes (24–26). Hence, these two CpG islands pinpoint the positions of potential candidate genes responsible for HD. They also demonstrate how remarkably close to the ends of a mammalian chromosome genes can be.

Given the relatively small amount of DNA between locus D4S90 and the telomere, the search for the HD gene must be entering its final stages. It would be quite unlikely for this region to contain more than 10 genes and it is very likely that the region contains far fewer than this. What is needed now is further characterization of the distal 325 kb of the chro-

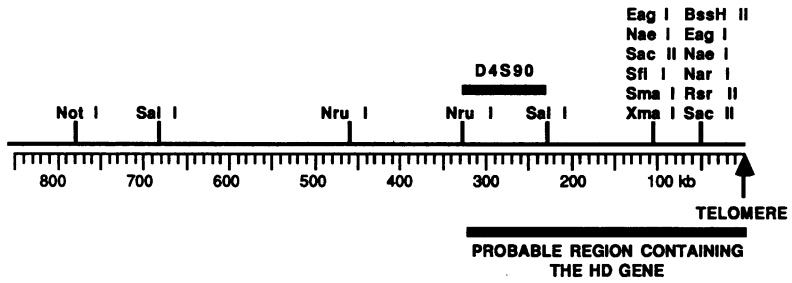

FIG. 4. Restriction map of the telomeric *Not* I fragment of the p arm of chromosome 4. The map is based on the hybridization results in Fig. 3, and on *Nru* I and *Sal* I partial digest experiments. The horizontal bar below locus D4S90 indicates the region to which the marker is confined. The HD locus has been reported to be distal to locus D4S90 (14).

mosome to discover any additional informative polymorphisms and any evolutionarily conserved or transcribed DNA sequences.

We are grateful to Dr. Sandra Youngman for supplying the D5 probe and to Dr. John Wasmuth for providing us with the HHW693, HHW416, and UCW56 cell lines. This research was supported by the National Cancer Institute (Grant CA39782), the U.S. Department of Energy (Grant DE-FG02-87ER-GD852), and the Hereditary Disease Foundation.

1. Hayden, M. R. (1981) *Huntington's Chorea* (Springer, New York).
2. Martin, J. B. & Gusella, J. F. (1986) *N. Engl. J. Med.* **315**, 1267–1276.
3. Wexler, N. S., Young, A. B., Tanzi, R. E., Travers, H., Starosta-Rubinstein, S., Penney, J. B., Snodgrass, S. R., Shoulson, I., Gomez, F., Arroyo, M. A. R., Penchaszadeh, G. K., Moreno, H., Gibbons, K., Faryniarz, A., Hobbs, W., Anderson, M. A., Bonilla, E., Conneally, P. M. & Gusella, J. F. (1987) *Nature (London)* **326**, 194–197.
4. Gusella, J. F., Wexler, N. S., Conneally, P. M., Naylor, S. L., Anderson, M. A., Tanzi, R. E., Watkins, P. C., Ottina, K., Wallace, M. R., Sakaguchi, A. Y., Young, A. B., Shoulson, I., Bonilla, E. & Martin, J. B. (1983) *Nature (London)* **306**, 234–238.
5. MacDonald, M. E., Anderson, M. A., Gilliam, T. C., Tranebjaerg, L., Carpenter, N. J., Magenis, E., Hayden, M. R., Healey, S. T., Bonner, T. I. & Gusella, J. F. (1987) *Genomics* **1**, 29–34.
6. Gilliam, T. C., Tanzi, R. E., Haines, J. L., Bonner, T. I., Faryniarz, A. G., Hobbs, W. J., MacDonald, M. E., Cheng, S. V., Folstein, S. E., Conneally, P. M., Wexler, N. S. & Gusella, J. F. (1987) *Cell* **50**, 565–571.
7. Gilliam, T. C., Bucan, M., MacDonald, M. E., Zimmer, M., Haines, J. L., Cheng, S. V., Pohl, T. M., Meyers, R. H., Whaley, W. L., Allitto, B. A., Faryniarz, A., Wasmuth, J. J., Frischauf, A.-M., Conneally, P. M., Lehrach, H. & Gusella, J. F. (1987) *Science* **238**, 950–952.
8. Wasmuth, J. J., Hewitt, J., Smith, B., Allard, D., Haines, J. L., Skarecky, D., Partlow, E. & Hayden, M. R. (1988) *Nature (London)* **332**, 734–736.
9. Hayden, M. R., Hewitt, J., Wasmuth, J. J., Kastelein, J. J., Langlois, S., Conneally, M., Haines, J., Smith, B., Hilbert, C. & Allard, D. (1988) *Am. J. Hum. Genet.* **42**, 125–131.
10. Smith, B., Skarecky, D., Bengtsson, U., Magenis, R. E., Carpenter, N. & Wasmuth, J. J. (1988) *Am. J. Hum. Genet.* **42**, 335–344.
11. Richards, J. E., Gilliam, T. C., Cole, J. L., Drumm, M. L., Wasmuth, J. J., Gusella, J. F. & Collins, F. S. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 6437–6441.
12. Whaley, W. L., Michiels, F., MacDonald, M. E., Romano, D., Zimmer, M., Smith, B., Leavitt, J., Bucan, M., Haines, J. L., Gilliam, T. C., Zehetner, G., Smith, C., Cantor, C. R., Frischauf, A.-M., Wasmuth, J. J., Lehrach, H. & Gusella, J. F. (1988) *Nucleic Acids Res.* **16**, 11769–11780.
13. Pohl, T. M., Zimmer, M., MacDonald, M. E., Smith, B., Bucan, M., Poustka, A., Volinia, S., Searle, S., Zehetner, G., Wasmuth, J. J., Gusella, J., Lehrach, H. & Frischauf, A.-M. (1988) *Nucleic Acids Res.* **16**, 9185–9198.
14. Robbins, C., Theilmann, J., Youngman, S., Haines, J., Altherr, M. J., Harper, P. S., Payne, C., Junker, A., Wasmuth, J. & Hayden, M. R. (1989) *Am. J. Hum. Genet.* **44**, 422–425.
15. Wasmuth, J. J., Carlock, L. R., Smith, B. & Immken, L. L. (1986) *Am. J. Hum. Genet.* **39**, 397–403.
16. Carlock, L. R., Smith, D. & Wasmuth, J. J. (1986) *Somatic Cell Mol. Genet.* **12**, 163–174.
17. Smith, C. L., Klco, S. R. & Cantor, C. R. (1988) in *Genome Analysis: A Practical Approach*, ed. Davies, K. (IRL, Oxford), pp. 41–71.
18. Smith, C. L., Warburton, P. E., Gaal, A. & Cantor, C. R. (1986) in *Genetic Engineering*, eds. Setlow, J. K. & Hollaneder, A. (Plenum, New York), Vol. 8, pp. 45–70.
19. Mathew, M. K., Smith, C. L. & Cantor, C. R. (1988) *Biochemistry* **27**, 9204–9210.
20. Smith, C. L. & Cantor, C. R. (1987) *Methods Enzymol.* **155**, 449–467.
21. Feinberg, A. P. & Vogelstein, B. (1983) *Anal. Biochem.* **132**, 6–13.
22. Cheng, J.-F., Smith, C. L. & Cantor, C. R. (1989) *Nucleic Acids Res.* **17**, 6109–6127.
23. Schwartz, D. C. & Cantor, C. R. (1984) *Cell* **37**, 67–75.
24. Bird, A., Taggart, M., Frommer, M., Miller, O. J. & Macleod, D. (1985) *Cell* **40**, 91–99.
25. Bird, A. P. (1986) *Nature (London)* **321**, 209–213.
26. Gardiner-Garden, M. & Frommer, M. (1987) *J. Mol. Biol.* **196**, 261–282.