Am. J. Hum. Genet. 43:854–859, 1988

New Approach for Isolation of VNTR Markers

Yusuke Nakamura, Mary Carlson, Karen Krapcho, Masao Kanamori, and Ray White

Howard Hughes Medical Institute, University of Utah Health Sciences Center, Salt Lake City

Summary

Elsewhere we have reported an efficient method for isolating VNTR (Variable Number of Tandem Repeats) markers. Several of the VNTR markers isolated in those experiments were sequenced, and a DNA sequence of 9 bp (GNNGTGGG) emerged as an apparent consensus sequence for VNTR markers. To confirm this result and to develop more VNTR markers, we synthesized nine different 18-base-long oligonucleotides whose sequences each included GNNGTGGG. When 102 cosmid clones selected by these oligonucleotides were tested for polymorphism, 34 (33%) of them showed multiallelic VNTR polymorphisms (average heterozygosity 68%). This procedure represents a new and efficient approach for isolating additional VNTR markers and supports the idea that the GNNGTGGG sequence may play an important role in the generation of the multiallelic systems within the human genome.

Introduction

The first multiallelic, highly polymorphic DNA marker was described by Wyman and White (1980). Following this, the insulin gene (Bell et al. 1982), the H-ras gene (Capon et al. 1983), and the zetaglobin pseudogene (Goodbourn et al. 1983) were each found to reveal multiallelic polymorphism due to the difference of the number of tandemly repeated short DNA sequences associated with these loci. As markers based on a variable number of tandem repeats (VNTRs) show a high heterozygosity within the population, they are more informative than diallelic systems for the construction of linkage maps of human chromosomes and for linkage studies to locate the defective locus in families segregating genetic diseases. Furthermore, their high degree of heterozygosity makes VNTRs markers a powerful tool for revealing reduction to homozygosity at loci that may be involved in carcinogenesis.

Jeffreys et al. (1985) reported "minisatellite" DNA sequences derived from the myoglobin gene, which showed a complex and highly polymorphic pattern on Southern blots of genomic DNA. This result implied that many highly polymorphic VNTR loci exist in the human genome, with the polymorphic systems sharing some sequence homology. On the basis of Jeffreys' finding, we constructed several oligonucleotides, each derived from a known VNTR sequence, to screen a human genomic cosmid library. At conditions of high stringency, we were able to isolate highly informative, locus-specific DNA markers for linkage studies. Of 300 cosmid clones that showed a positive signal with an oligonucleotide, 77 (26%) revealed VNTR polymorphism; we identified GNNGTGGG as a sequence common to 10 of the VNTR markers (Nakamura et al. 1987). Wong et al. (1987) reported the core sequence of several highly polymorphic loci as GGAGGTGGG-CAGGAR (A or G)G, which also includes GNNGTGGG.

To identify a wider range of VNTR loci, we synthesized several different oligonucleotides containing the sequence GNNGTGGG and used each of them to probe a genomic DNA library. This technique constitutes a new and productive approach to the efficient isolation of VNTR markers.

Material and Methods

DNA and Restriction Enzymes

DNA was isolated from lymphoblasts according to the method described by Bell et al. (1981). Restriction enzymes were purchased from Molecular Biology Resources, Milwaukee. Oligonucleotides were synthesized and provided to us by Dr. R. Gesteland. Cosmid libraries were gifts from Dr. Y. W. Kan (Lau and Kan 1983).

Received June 21, 1988; revision received August 2, 1988.

Address for correspondence and reprints: Yusuke Nakamura, M.D., Howard Hughes Medical Institute, 603 Wintrobe Building, University of Utah Health Sciences Center, Salt Lake City, UT 84132.

0002-9297/88/4306-0007$02.00

Southern Hybridizations

Five micrograms of lymphoblast DNA from each of six unrelated individuals was digested with restriction enzymes (*Msp*I, *Taq*I, *Rsa*I, *Bgl*II, *Pst*I, or *Pvu*II). DNAs were electrophoresed in agarose gels and transferred to nylon membrane filters (Gelman) by the sodium hydroxide method (Reed and Mann 1985). Prehybridizations were carried out in a solution containing 50% formamide, 5 × SSC (1 × SSC = 0.15 M NaCl/0.015 M sodium citrate), 10 × Denhardt solution (1 × Denhardt = 0.02% polyvinylpyrrolidone/0.02% bovine albumin/0.02% Ficoll-400), 0.5% SDS, 50 mM $NaPO_4$ (pH 6.7), and 250 μg human placental DNA/ml at 42 C for more than 24 h. Probe DNAs were labeled with [^{32}P]dCTP by the hexanucleotide priming method (Feinberg and Vogelstein 1984). Hybridizations were carried out under the same conditions, except that the solution contained 10% dextran sulfate instead of 10 × Denhardt's solution. Filters were washed in 2 × SSC and 0.1% SDS at room temperature for 5 min and twice in 0.1 × SSC and 0.1% SDS at 65 C for 15 min.

Chromosomal Assignment

The new VNTR markers were genotyped in 59 3-generation reference families that included 40 from the CEPH (Centre d'Etude du Polymorphisme Humain, Paris) panel. Genotypes were analyzed with the GMS linkage program (Lathrop et al. 1988); chromosomal assignments for 18 of the new markers were made according to linkage with markers whose chromosomal locations had already been mapped. Only one among the newly mapped clones (cMCOB19) defined a locus that had been reported previously (D19S20, defined by pJCZ3.1; Nakamura et al. 1987). The other 17 clones identified a completely new set of VNTR marker loci.

Results

"Families" of VNTRs

To examine the feasibility of using a known VNTR core sequence for isolating related VNTR loci, we probed genomic Southern blots with an 18-base-long oligonucleotide (GGAGCAGTGGGNNNTACA) that is part of the 30-base consensus sequence in the repeating unit at the pYNH24 (D2S44) locus (Nakamura et al. 1987). The results are shown in figure 1*a*. The bands identified as locus YNH24 by comparison with a genomic blot made with the plasmid as probe (fig. 1*b*) were more intense than the extra bands, probably owing to slight consensus-sequence differences among the loci represented by the fainter bands. Additionally, two or more highly polymorphic systems were evident. We then screened a human cosmid library with the 18-base oligonucleotide, under the same conditions used for the genomic Southern blotting. From the one-genome-equivalent library (75,000 colonies) we isolated five positive clones. Three of the five positive cosmids (YNA4, YNA12, and YNA13 in table 3), tested for polymorphism with three restriction enzymes (*Msp*I, *Taq*I,

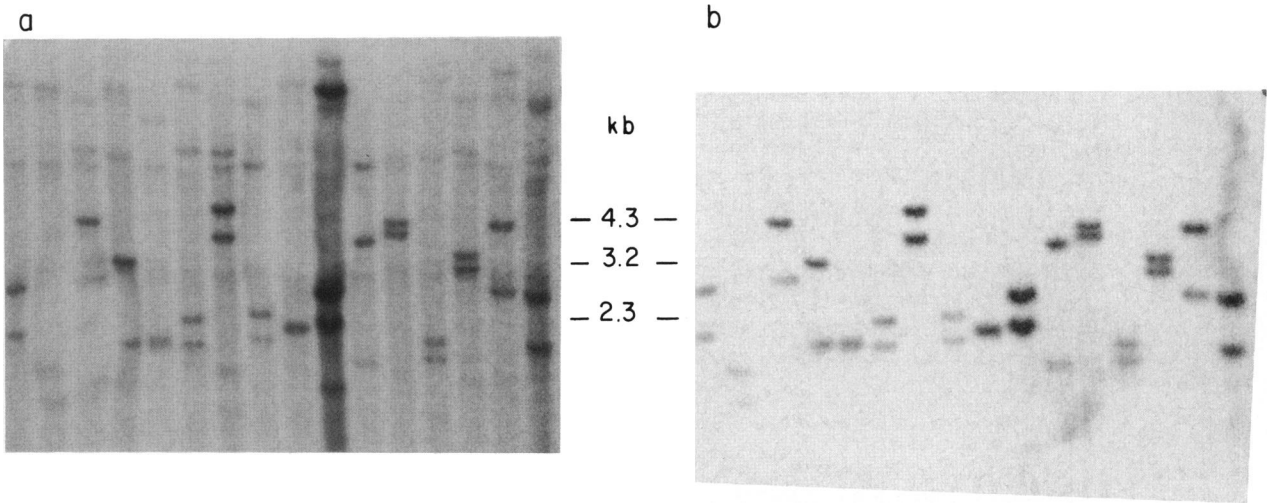

Figure 1 Southern blots of genomic DNA from 16 unrelated individuals, hybridized with (*a*) the 18-base-long oligonucleotide that is part of the consensus sequence of pYNH24 (GGAGCAGTGGGNNNTACA) and (*b*) whole-plasmid pYNH24, under the conditions described in Material and Methods for plasmids and in table 1 for oligonucleotides.

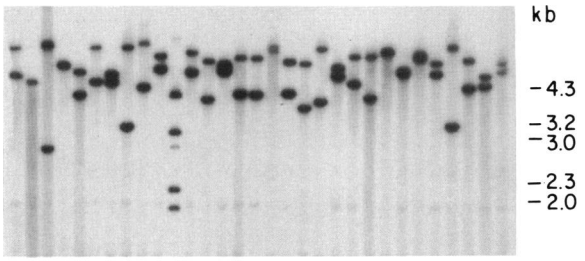

Figure 2 Southern blot of genomic DNA from 30 unrelated individuals, digested with *MspI* and probed with whole-cosmid DNA of YNA13.

and *PstI*), presented VNTR polymorphisms that were different from the pYNH24 locus. Genomic Southern blots probed with cYNA13 revealed 97% heterozygosity among 100 unrelated Caucasians (fig. 2 shows a blot for 30 of these individuals); this locus was mapped by linkage to markers near the telomere of the long (q) arm of chromosome 1 (O'Connell et al., in press).

Screening of Genomic Cosmid Library with Oligonucleotides

Extending this approach, we synthesized nine different 18-base-long oligonucleotides for screening a human genomic cosmid library. Both the sequence of each oligonucleotide and the screening conditions used are shown in table 1. All oligonucleotides included "GNNGTGGG"; the 12-base-long sequences outside this core sequence were chosen randomly. Each oligonucleotide was a mixture of 256–1,024 different sequences, because each included four or five N's (N = A, G, C, or T). Screening of a one-genome-equivalent cosmid library (almost 75,000 colonies) yielded 3–35 positive clones for each oligonucleotide.

Table 2

Summary of VNTR Screening

Oligonucleotide	No. of Cosmid Clones Tested	No. of VNTR Markers Isolated
YN24	5	3
YN54	8	4
YN55	8	3
YN56	14	5
YN58	8	3
YN59	3	1
YN63	27	8
YN69	9	1
YN73	4	1
YN75	16	5
Total	102	34 (33% of total)

Testing for Polymorphism

Table 2 shows the results of screening for polymor-

Table 1

Sequences of 13 Oligonucleotides and Hybridization Conditions for a Cosmid Library

Oligonucleotides	Hybridization Temperature (°C)	Washing Temperature (°C)
YN24 (GGAGCAGTGGGNNNTACA)	55	58
YN54 (GAGGGTGGNGGNTCTNNG)	55	60
YN55 (GGGAGGAGTGGGNNNNNG)	55	60
YN56 (GGGCGGGGTGGGNNNNNG)	55	65
YN58 (GGGNGTGGGGGNGTNNNG)	55	63
YN59 (GGGNGTGGGNNNNCACAC)	55	60
YN63 (GGGTGGGGTGGGNNNNNG)	55	63
YN71 (GNNNNGGTGGGGCCGGTG)	55	65
YN73 (CCCGTGGGGCCGNNNNGG)	55	67
YN75 (GTGCCGTGGGTGCNNNNG)	55	63

NOTE.—Hybridization was done in a solution containing 50 mM Tris HCl (pH 7.4), 10 μg yeast tRNA/ml, 1 × Denhardt's solution, and 6 × SSC. Filters were washed in 5 × SSC, 0.1% SDS at room temperature for 5 min and subsequently twice in 5 × SSC, 0.1% SDS for 5 min at the temperature shown in the table.

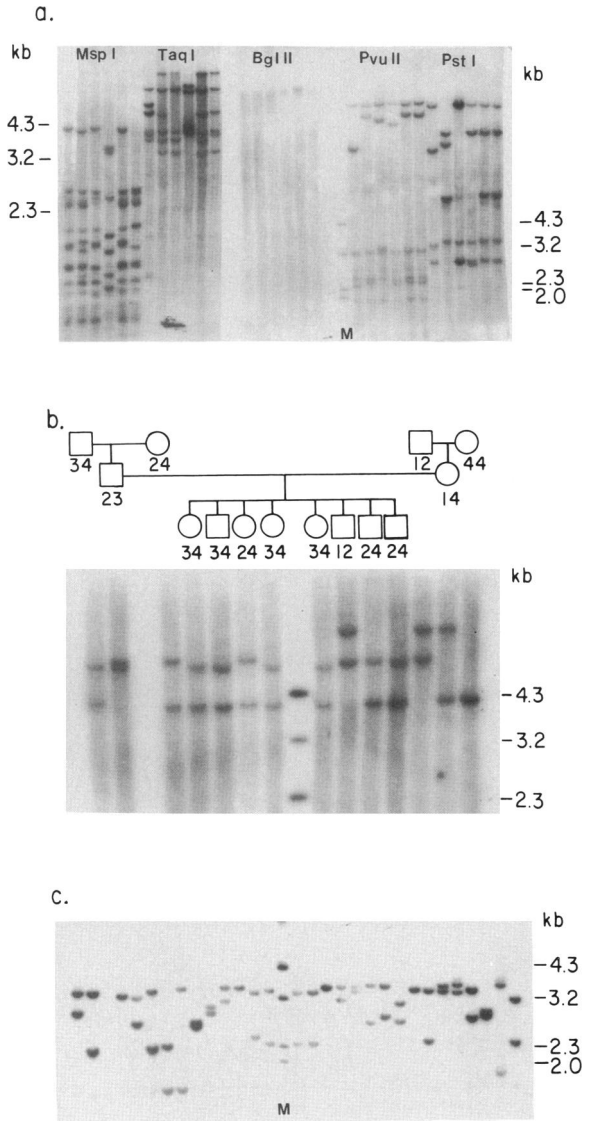

Figure 3 *a*, Southern blots of genomic DNA digested with five different enzymes and hybridized with cosmid MCOD13, from six unrelated individuals. The size markers on the left are for *Msp*I and *Taq*I, and those on the right are for *Bgl*II, *Pvu*II, and *Pst*I; *b*, Southern blots of genomic DNA digested with *Taq*I and hybridized with KKA25, from 14 individuals in kindred 1444; *c*, Southern blots of genomic DNA digested with *Hin*fI and hybridized with KKA39, from 29 unrelated individuals.

phism among the positive clones. One hundred two cosmid clones (including the YNH24 "family") were tested for polymorphisms by using the DNAs from six unrelated individuals and the six restriction enzymes listed in Material and Methods.

Of 102 cosmid clones tested, 34 clones (33%) showed single-locus VNTR polymorphisms. For example, cosmid MCOD13 revealed the VNTR type of polymorphism among six unrelated DNAs digested with *Msp*I, *Taq*I, *Pst*I, or *Pvu*II (fig. 3*a*). *Bgl*II did not detect VNTR polymorphism at this locus, either because the sequence of the repeating unit includes the recognition site for this enzyme or because the polymorphic alleles are so small that they run out from the gel. Figure 3*b* shows the cosegregation pattern of alleles at the KKA25 locus in kindred 1444; the grandparental source of each allele is clear in all the children. With *Hin*fI, cosmid KKA39 (fig. 3*c*) showed five alleles among 29 unrelated individuals; of the 29, 24 (83%) were heterozygotes.

We also tested for polymorphism two 18-base-long oligonucleotides that do not include the "GNNGTGGG" sequence. The first of these, GATAGATAGATANNNNGA, identified 14 cosmid clones; none of them revealed a VNTR pattern of polymorphism. The other, TTTTATAATTAAAATA, reflected the core sequence of the 3′ VNTR region of the apolipoprotein B gene (Knott et al. 1986); of 25 cosmid clones isolated by this oligonucleotide, only one revealed VNTR polymorphism (data not shown).

Chromosomal Localization

Eighteen of the 34 VNTR markers listed in table 3 were genotyped in 59 3-generation families and were assigned chromosomal localizations by linkage analysis against DNA markers mapped previously. The average heterozygosity of these markers was 68% (71% for markers with more than three alleles), which is nearly the same as the heterozygosity found among the sets of VNTR markers reported elsewhere (Nakamura et al. 1987). The new VNTRs are scattered on many chromosomes, but 60% of them have been mapped near telomeres (Y. Nakamura and R. White, unpublished data).

Discussion

We have developed a new approach to the isolation of VNTR markers for the human genetic linkage map, on the basis of hybridization of genomic DNA with oligonucleotide probes that incorporate the sequence GNNGTGGG. For example, the 18-base consensus sequence of pYNH24 (D2S44), which contains GCAGTGGG, allowed us to isolate three highly informative VNTR markers with an average heterozygosity of 85%. When we screened a cosmid library with 10 different GNNGTGGG-containing oligonucleotides, the num-

Table 3

Characterization of VNTR Markers

Probe	Enzyme[a]	Allele Size Range (kb)	No. of Alleles	Heterozygosity (%)	Chromosome	Oligonucleotide
YNA4	*Msp*I	5.0–10.0	>10	85[b]	2	YN24
YNA12	*Msp*I	2.5–3.0	6	63[b]	13	YN24
YNA13	*Msp*I	2.0–10.0	>20	97[b]	1	YN24
MCOA5	*Msp*I	2.7/2.8	2	40[c]	ND	YN59
MCOA7	*Eco*RI	4.9–5.5	5	80[c]	ND	YN55
MCOA9	*Pvu*II	3.3–6.0	8	83[c]	ND	YN55
MCOA12	*Taq*I	3.5–7.6	3	49[b]	9	YN55
MCOB5	*Pst*I	5.3–8.0	5	70[b]	19	YN56
MCOB12	*Taq*I	2.4–4.3	4	78[b]	6	YN56
MCOB17	*Pst*I	1.2–2.1	5	83[c]	ND	YN56
MCOB19	*Pst*I	3.5–5.5	5	75[b]	19	YN56
MCOB23	*Pst*I	7.3/7.6	2	50[c]	ND	YN56
MCOC6	*Pst*I	3.5–4.3	6	67[c]	ND	YN54
MCOC12	*Msp*I	.8–2.0	7	60[b]	14	YN54
MCOC14	*Pst*I	3.0–3.5	4	67[b]	4	YN58
MCOC35	*Msp*I	3.2–4.2	6	75[c]	ND	YN58
MCOC36	*Taq*I	5.0/5.5	2	46[b]	16	YN54
MCOC37	*Bgl*II	6.0–8.0	6	75[b]	ND	YN58
MCOC46	*Pst*I	1.5–3.0	4	67[b]	13	YN54
MCOD13	*Pvu*II	4.0–10.0	8	85[b]	3	YN71
MCOE11	*Msp*I	1.2/1.6	2	50[c]	ND	YN73
MCOE13	*Hae*III	2.2–2.7	4	67[c]	ND	YN75
MCOE32	*Msp*I	1.0–2.0	5	78[b]	2	YN75
MCOE34	*Msp*I	3.3–4.5	4	75[c]	ND	YN75
MCOE41	*Msp*I	1.2–2.0	6	83[c]	ND	YN75
MCOE51	*Msp*I	.6–2.1	3	60[c]	ND	YN75
KKA12	*Msp*I	2.0–2.8	6	65[b]	7	YN63
KKA17	*Pvu*II	2.5/2.7	2	45[c]	ND	YN63
KKA22	*Taq*I	2.5–3.2	3	60[b]	16	YN63
KKA25	*Msp*I	3.8–6.0	7	75[b]	ND	YN63
KKA28	*Pst*I	2.5–4.5	8	83[c]	ND	YN63
KKA35	*Msp*I	1.8–2.0	3	37[b]	17	YN63
KKA39	*Msp*I	2.0–4.0	>10	83[b]	14	YN63
KKA40	*Pvu*II	1.5–1.8	3	44[b]	9	YN63

NOTE.—ND = not yet determined.

[a] Only the enzymes that gave the best resolution are shown.

[b] Result for 120 unrelated individuals.

[c] Result for 18 unrelated individuals.

ber of positive clones varied but there was no significant difference in the proportion of VNTR markers.

This study shows that oligonucleotides incorporating GNNGTGGG are able to enrich a cosmid library for a VNTR locus 10- to 20-fold (33% of clones showing VNTR polymorphism) in comparison with screening with random probes (2%–3%) or with oligonucleotides that do not include GNNGTGGG.

The high degree of heterozygosity observed among the YNH24 family of markers suggests that this consensus sequence is highly recombinogenic and raises the possibility that sequences additional to GNNGTGGG might play an important role in the generation of highly polymorphic VNTR loci. However, there is no homology outside the GNNGTGGG sequence between the YNH24 family of markers and the hypervariable loci sequenced by Wong et al. (1987).

The fact that the consensus sequence GNNGTGGG bears some homology to the *chi* sequence of lambda phage (GCTGTGG) supports the notion that the VNTR core sequence is a hot spot for recombination. However, other observations suggest a separate function of

the VNTR sequence: (1) the VNTR sequence at the 3′ flanking region of the H-*ras* gene has a function in transcriptional control (Spandidos and Holmes 1987); (2) the consensus sequence of many enhancer sequences (GNTGTGG$^{TTT}_{AAA}$) is similar to GNNGTGGG (e.g., see Ondek et al. 1988); and (3) VNTR sequences exist in a flanking region or in an intron of certain genes (e.g., insulin, H-*ras*, and retinoblastoma), where they might have a role in control of gene expression.

The chromosomal localizations of 18 of the new VNTR markers, determined by linkage analysis, add to the body of data showing that VNTR loci are scattered on many chromosomes and that there seems to be a bias toward location near telomeric regions (Donis-Keller et al. 1987; Kondoleon et al. 1987; Nakamura et al. 1988*a*, 1988*b*; Royle et al. 1988).

Acknowledgments

We are grateful to Drs. Jean-Marc Lalouel, Mark Lathrop, Mark Leppert, and Peter O'Connell for valuable discussions. We also thank Linda Ballard, Janice Gill, Mary Hadley, Leslie Jerominski, and Lesa Nelson for technical assistance and Ruth Foltz for editorial work in the preparation of the manuscript. R. White is an investigator for the Howard Hughes Medical Institute.

References

Bell, G. I., J. Karam, and W. Rutter. 1981. Polymorphic DNA region adjacent to the 5′ end of the human insulin gene. Proc. Natl. Acad. Sci. USA **78**:5759–5763.

Bell, G. I., M. J. Selby, and W. J. Rutter. 1982. The highly polymorphic region near the human insulin gene is composed of simple tandemly repeating sequences. Nature **295**:31–35.

Capon, D. J., E. Y. Chen, A. D. Levinson, P. H. Seeburg, and D. V. Goeddel. 1983. Complete nucleotide sequences of the T24 human bladder carcinoma oncogene and its normal homologue. Nature **302**:33–37.

Donis-Keller, H., and 32 other authors. 1987. A genetic linkage map of the human genome. Cell **51**:319–337.

Feinberg, A., and B. Vogelstein. 1984. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **137**:266–267.

Goodbourn, S. E. Y., D. R. Higgs, J. B. Clegg, and D. J. Weatherall. 1983. Molecular basis of length polymorphism in human ζ-globin gene complex. Proc. Natl. Acad. Sci. USA **80**:5022–5026.

Jeffreys, A. J., V. Wilson, and S. L. Thein. 1985. Hypervariable "mini-satellite" regions in human DNA. Nature **314**: 67–73.

Knott, T. J., S. C. Wallis, R. J. Pease, L. M. Powell, and J. Scott. 1986. A hypervariable region 3′ to the human apolipoprotein B gene. Nucleic Acids Res. **14**:9215.

Kondoleon, S., H. Vissing, X. Luo, R. Magenis, J. Kellogg, and M. Litt. 1987. A hypervariable RFLP on chromosome 17p13 is defined by an arbitrary single copy probe p144-D. Nucleic. Acids Res. **15**:10605.

Lathrop, M., Y. Nakamura, P. Cartwright, P. O'Connell, M. Leppert, C. Jones, H. Tateishi, T. Bragg, J.-M. Lalouel, and R. White. 1988. A primary genetic map of markers for human chromosome 10. Genomics **2**:157–164.

Lau, Y.-F., and Y. W. Kan. 1983. Versatile cosmid vectors for the isolation, expression, and rescue of gene sequences: studies with the human alpha-globin gene cluster. Proc. Natl. Acad. Sci. USA **80**:5225–5229.

Nakamura, Y., M. Lathrop, P. O'Connell, M. Leppert, D. Barker, W. Wright, M. Skolnick, S. Kondoleon, M. Litt, J.-M. Lalouel, and R. White. 1988*a*. A mapped set of DNA markers for human chromosome 17. Genomics **2**:302–309.

Nakamura, Y., M. Lathrop, P. O'Connell, M. Leppert, J.-M. Lalouel, and R. White. 1988*b*. A primary map of ten DNA markers and two serological markers for human chromosome 19. Genomics **3**:67–71.

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 **235**:1616–1622.

O'Connell, P., M. Lathrop, Y. Nakamura, M. Leppert, R. H. Ardinger, J. L. Murray, J.-M. Lalouel, and R. White. Twenty-eight loci form a continuous linkage map of markers for human chromosome 1. Genomics (in press).

Ondek, B., L. Gloss, and W. Herr. 1988. The SV40 enhancer contains two distinct levels of organization. Nature **333**: 40–45.

Reed, K. C., and D. A. Mann. 1985. Rapid transfer of DNA from agarose gels to nylon membranes. Nucleic Acids Res. **13**:7207–7221.

Royle, N. J., R. Clarkson, Z. Wong, and A. J. Jeffreys. 1988. Preferential localization of hypervariable minisatellites near human telomeres. Cytogenet. Cell Genet. **46**:685.

Spandidos, D. A., and L. Holmes. 1987. Transcriptional enhancer activity in the variable tandem repeat DNA sequence downstream of the human Ha-*ras*I. FEBS Lett. **218**:41–46.

Wong, Z., V. Wilson, I. Patel, S. Povey, and A. J. Jeffreys. 1987. Characterization of a panel of highly variable minisatellites cloned from human DNA. Ann. Hum. Genet. **51**:269–288.

Wyman, A. R., and R. White. 1980. A highly polymorphic locus in human DNA. Proc. Natl. Acad. Sci. USA **77**: 6754–6758.