Jpn. J. Human Genet. **31**, 249–258, 1986

ISOLATION OF DNA CLONES REVEALING RESTRICTION FRAGMENT LENGTH POLYMORPHISMS IN THE HUMAN GENOME

Isamu NISHISHO,[1] Tetsuro MIKI,[2,3] Hideo TATEISHI,[1]
Shin-ichiro TAKAI,[1] Kazuyoshi MOTOMURA,[1] Jun NAKURA,[2]
Yuichi KUMAHARA,[2] Takesada MORI,[1] and Tasuku HONJO[3,4]

[1]*Second Department of Surgery,* [2]*Department of Medicine and Geriatrics,*
[3]*Department of Genetics, Osaka University Medical School,*
Osaka 553, Japan
[4]*Department of Medical Chemistry, Kyoto University*
Faculty of Medicine, Kyoto 606, Japan

Summary A recombinant human DNA library was constructed using pUC 18 as the cloning vector. Plasmid DNA isolated from a small scale culture was used as the hybridization probe; many recombinant clones could be easily tested for their ability to detect restriction fragment length polymorphisms (RFLPs). Forty-five arbitrary single copy DNA fragments were isolated from this library and five clones revealed RFLPs. Probe OS-5 had three alleles and probe OS-7, which detected insertion/deletion polymorphisms, had several alleles.

Apart from these clones, two polymorphic DNA fragments were isolated from the pBR322 plasmid library and another two from the Charon 4A phage library. Although only four restriction enzymes were employed to detect polymorphisms, the efficiency of detecting polymorphisms was reasonable.

These nine clones will serve as useful markers for linkage studies.

INTRODUCTION

Polymorphic DNA markers, which detect restriction fragment length polymorphisms (RFLPs), are useful in making a human gene map. In addition to cloned DNA segments which can represent genes of known specificity, anonymous fragments isolated from a DNA library can be used to detect DNA polymorphisms. By means of recombinant DNA techniques, the number of such genetic markers available for linkage studies has been enlarged. For construction of a complete

Received April 29, 1986; Accepted May 19, 1986

RFLPs linkage map, at least 150 loci would be needed if spaced evenly no more than 20 cM apart (Botstein *et al.*, 1980). However, to obtain evenly spaced loci, a larger number of marker loci must be placed on the map.

As most human genetic diseases have no specific phenotype in cultured cells, there is no clue to enable us to determine the chromosomal localization of their genes using hybrid cell mapping approaches. Linkage analysis is the only method applicable for determination of their loci on the gene map. By using arbitrary RFLPs for linkage analysis, Gusella *et al.* (1983) demonstrated that the G8 probe linked to the locus of Huntington's disease. Recently Reeders *et al.* (1985) used the α-globin gene to determine the locus of adult polycystic kidney disease and three researchers demonstrated tight linkage between cystic fibrosis locus and polymorphic DNA markers on the long arm of chromosome 7 (Knowlton *et al.*, 1985; White *et al.*, 1985; Wainwright *et al.*, 1985).

In order to determine the loci of multiple endocrine neoplasia type 2A (MEN 2A) and familial polyposis coli (FPC), both of which are inherited as autosomal dominant traits, nine polymorphic DNA fragments derived from a genomic DNA library were cloned. Taking advantage of the fact that the copy number of pUC 18 in the host cell is high, we used this plasmid as the cloning vector. The quantity of plasmid DNA isolated from 5 ml LB culture was enough to use as a hybridization probe to detect polymorphisms. One of the cloned polymorphic DNA fragments detected three alleles and another one indicated insertion/deletion polymorphisms. These two clones are useful markers for linkage analysis of genetic diseases whose loci have not been localized to specific chromosomes.

MATERIALS AND METHODS

Library construction

1) pUC 18 library: High molecular weight DNA was isolated from human leukocytes and digested completely with *Hin*dIII. DNA fragments with molecular size 2–8 kb were collected by agarose gel electrophoresis. These fragments were ligated into *Hin*dIII digested pUC 18. The resulting ligation products were used to transform *E. coli* strain HB101, then they were plated on LB plates containing ampicillin. Transformants grown on nitrocellulose filters according to the method of Hanahan and Meselson (1980) were probed with nick-translated total human DNA. Bacteria containing plasmids carrying no repetitive sequence were tooth-picked into 5 ml of liquid LB medium. Plasmids were isolated from 5 ml culture by alkaline method and small aliquots of each plasmid were digested with *Hin*dIII. *Hin*dIII digested plasmids were electrophoresed on 0.7% agarose gels and transferred to nitrocellulose by the method of Southern (1975).

Total human DNA, radio-labeled by nick-translation, was hybridized to the plasmid DNA on the filters. Because plasmids carrying human repetitive sequences

yielded signals, they could be excluded. Plasmids that contained single copy DNA fragment were used as probes that detect RFLPs.

2) pBR 322 library: *Eco*RI-*Hin*dIII digested human DNA fragments were ligated into pBR 322, according to the method of Barker *et al.* (1984).

3) Charon 4A library: The Charon 4A *Eco*RI library was kindly provided by T. Maniatis. Subcloned small segments from phages were used as hybridization probes.

DNA isolation

Human DNA was prepared from white blood cells donated by healthy Japanese volunteers. Heparinized human peripheral blood, usually 20 ml, was collected. The blood sample was mixed with equal volume of 3% (w/v) dextran made up in normal saline. The erythrocytes were allowed to sediment from the serum standing still for 30 min. The leukocyte-enriched supernatant was collected, centrifuged and the pellet washed once in normal saline. The pellet was resuspended in 3 ml of 1 × SET (0.15 M NaCl, 1 mM EDTA, 20 mM Tris-HCl, pH 7.8), and 135 μl of 10% sodium dodecyl sulphate (SDS) and 450 μl of 5 M $NaClO_4$ were added. After gentle stirring, the mixture was then extracted with phenol and several times with chloroform/isoamyl alcohol (24 : 1). DNA was dialyzed completely against 0.1 × SSC (0.15 M NaCl, 15 mM sodium citrate).

Restriction enzyme digestion and DNA blotting

DNA samples were digested with *Msp*I, *Taq*I, *Hin*dIII, and *Eco*RI using at least a 5-fold excess of enzyme under the manufacturer's recommended conditions. The resulting fragments were separated on 0.7% agarose gels, and after electrophoresis the DNA was transferred to nitrocellulose by the method of Southern. After transfer, filters were baked under vacuum at 80°C for 3 hr.

Nick translation

Whole plasmids isolated from the 5 ml culture by alkaline method were used as probes. Probes were labeled by nick-translation with [^{32}P]-alpha-dCTP (Amersham, England) to specific activities of at least 5×10^8 cpm/μg. Radio-active probes were hybridized to genomic blots at 65°C for 15–20 hr in 1 × Denhardt's solution, 1 M NaCl, 50 mM Tris-HCl, pH 7.4, 10 mM EDTA, 0.1% SDS and 0.1 mg/ml denatured sonicated salmon sperm DNA. Filters were washed once at room temperature in 2 × SSC and twice at 65°C in 0.1 × SSC, 0.1% SDS for 30 min. The filters were exposed to XRP-1 film (Kodak) backed by a Lightning Plus intensifying screen (DuPont) at −80°C for 1–5 days.

RESULTS

We obtained the nine polymorphic DNA clones listed in Table 1. Two were cloned from the pBR 322 plasmid library, two from the Charon 4A phage library and five from the pUC 18 plasmid library. Clones that detect polymorphism at

Table 1. List of polymorphic DNA clones.

Lab. Symbol	Clone			Alleles				
	Name	Vector	Size (kb)	Name	Enzyme	Length (kb)		Frequency
OS-1	P1-1	pBR322	1.5	A1	*Taq*I	5.2		0.34
				A2		3.6	1.6	0.66
OS-2	#118	Ch4A	4.0	A1	*Hin*dIII	7.4		0.38
				A2		3.8		0.62
				B1	*Taq*I	11.0		0.40
				B2		6.0		0.60
OS-3	#119	Ch4A	2.0	A1	*Taq*I	6.0		0.62
				A2		5.4		0.38
OS-4	#148	pBR322	1.0	A1	*Taq*I	8.6		0.42
				A2		6.7		0.58
OS-5	pUC18-3	pUC18	2.0	A1	*Msp*I	9.4		0.35
				A2		8.0		0.21
				A3		6.2		0.44
OS-6	pUC18-15	pUC18	3.0	A1	*Taq*I	10.1		0.17
				A2		8.0	6.7	0.83
OS-7	pUC18-27	pUC18	4.4		*Msp*I	7–10		Several
					*Taq*I	4–7		Several
					*Eco*RI	6–9		Several
					*Hin*dIII	4–7		Several
OS-8	pUC18-32	pUC18	4.0	A1	*Taq*I	10.1		0.42
				A2		9.0		0.58
OS-9	pUC18-43	pUC18	5.0	A1	*Taq*I	9.8		0.48
				A2		8.0		0.52

These clones were isolated from three different genomic libraries.

low frequency are not listed, because such clones are not useful markers when employed in linkage analysis of genetic diseases. In the case of cloning from pUC 18 library, 1,500 recombinant plasmids at a time were probed with nick-translated total human DNA. Eighty colonies yielded no signals and they were toothpicked into 5 ml of liquid LB medium. Radio-labeled total human DNA was hybridized to the plasmid DNA isolated from the 5 ml culture. After second hybridization, 20 plasmids that contained a single copy DNA insert were obtained and were used as probes to detect DNA polymorphisms. Each probe was hybridized to Southern transfers of DNA digested with *Msp*I, *Taq*I, *Hin*dIII, and *Eco*RI from twelve unrelated individuals. Four of 20 clones revealed RFLPs.

Figure 1 presents examples of RFLPs observed in DNA from twelve unrelated

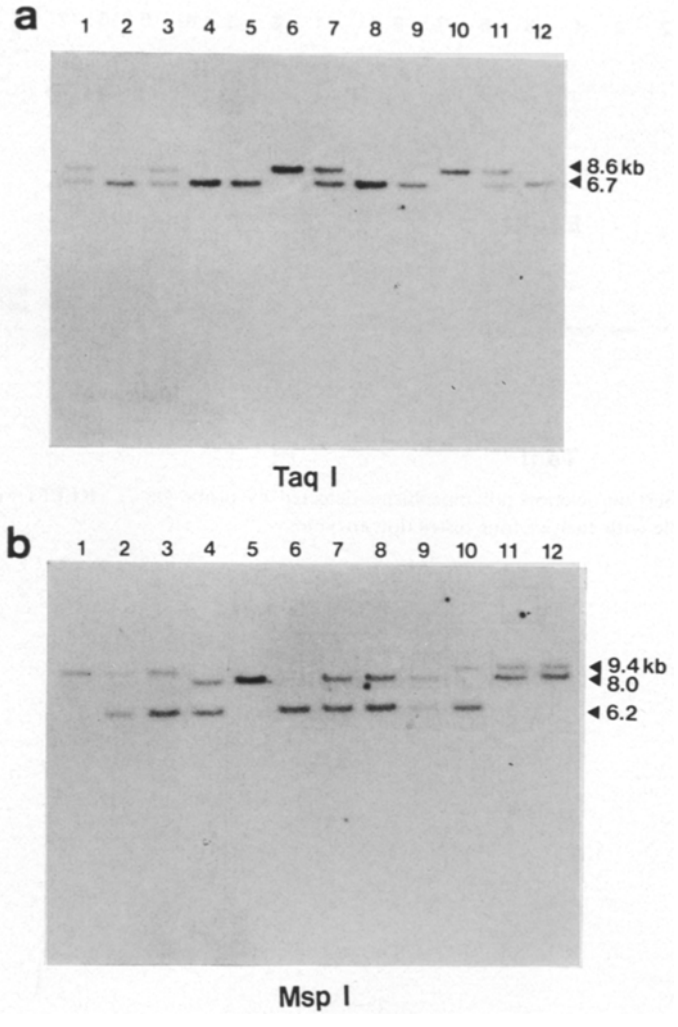

Fig. 1. Restriction fragment length polymorphisms detected by probe OS-4 and OS-5. a) The 1.0-kb fragment from #148 was used as a hybridization probe on Southern transfer of *Taq*I-digested DNA from twelve unrelated individuals. b) The 2.0-kb fragment from pUC 18-3 was used as a hybridization probe on Southern transfers of *Msp*I-digested DNA from twelve unrelated individuals.

individuals. Probe OS-4 cloned from the pBR 322 library detected two alleles of lengths 8.6 and 6.7 kb in *Taq*I digested human DNA. To determine the frequency of each allele, *Taq*I digested DNA from 36 unrelated Japanese were analyzed. The frequency of 8.6 kb allele was 0.42. Probe OS-5 cloned from the pUC 18 library detected three alleles of lengths 9.4, 8.0, and 6.2 kb in *Msp*I digested DNA. The frequency of each allele was 0.35, 0.21, and 0.44, respectively.

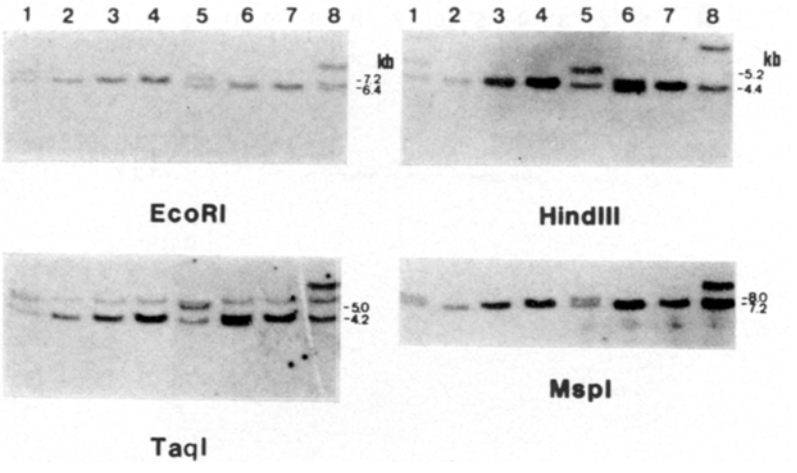

Fig. 2. Insertion/deletion polymorphisms detected by probe OS-7. RFLPs were detectable with each of four restriction enzymes.

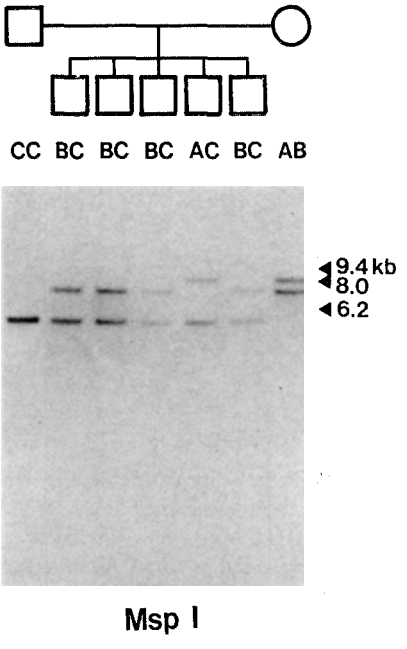

Fig. 3. Inheritance of polymorphic alleles detected by probe OS-5 in a small nuclear family.

Figure 2 presents the insertion/deletion DNA polymorphisms identified with probe OS-7. Several different alleles are detectable with each of four restriction enzymes.

To confirm the Mendelian inheritance of the RFLPs, its segregation was analyzed in several families. Figure 3 presents the hybridization pattern obtained with OS-5 for one of the families in this study. As expected, three alleles at this marker locus are inherited in Mendelian fashion.

DISCUSSION

In this study, nine polymorphic DNA fragments with lengths of 1–5 kb were cloned from three different genomic libraries. Three kinds of libraries have been used for cloning polymorphic DNA fragments: 1) a genomic library as we used in this report; 2) a genomic library prepared from a specific human chromosome or chromosome fragment, for instance a library from human-mouse hybrid cell DNA (Cavenee *et al.*, 1984) or from DNA of flow sorted metaphase chromosomes (Cooper *et al.*, 1985); 3) and a cDNA library of gene transcripts (Helentjaris and Gesteland, 1983). Using a library prepared from hybrid cell DNA or DNA of flow sorted metaphase chromosomes has the advantage of obtaining fragments from specific chromosomes. In the case of linkage analysis of genetic diseases whose loci have been already localized to specific chromosomes, such kind of library is very useful. But the loci of most human genetic diseases including MEN 2A and FPC have not been mapped to specific chromosomes. Therefore we chose a genomic library containing DNA fragments from all human chromosomes. Helentjaris and Gesteland (1983) used randomly isolated cDNA clones to detect DNA polymorphisms. cDNA clones representing members of gene families revealed many RFLPs but single-copy cDNA clones failed to detect RFLPs.

The probes have been divided into three groups according to their length: very large genomic fragments (35–45 kb) (Litt and White, 1985), large genomic fragments 12–20 kb) (Feder *et al.*, 1985), and smaller genomic fragments (1–3 kb) (Barker *et al.*, 1984). As the longer the probe segment, the more restriction enzyme recognition sites can be examined, it is very likely that long probes can detect multi-allelic loci. On the other hand, the longer the probe segment, the more the recombinant clones have repeptitive sequences. We cloned two polymorphic fragments from Charon 4A phage library containing large genomic fragments. Laborious processess of subcloning small segments from phages were needed. Next we chose pUC 18 as the cloning vector. Because of its high copy number in the host cell, a 5 ml culture of HB101 yielded enough quantity of plasmid DNA for use as a hybridization probe. By means of this labor-saving method, many recombinant clones could be checked for their ability to detect RFLPs.

We employed only four restriction enzymes, *Msp*I, *Taq*I, *Hin*dIII, and *Eco*RI to detect polymorphisms. According to the report of Barker *et al.* (1984), FRLPs are detected at a higher frequency in restriction enzyme recognition sites that contain the CpG dimer. In the human genome, most of these dinucleotides are methylated (Van der Ploeg and Flavell, 1980; Cooper, 1983) and 5-methyl-cytosine is

frequently replaced by thymidine due to the deamination of the methylated base (Vogel, 1972; Vogel and Kopun, 1977). That is, both *Msp*I(CCGG) and *Taq*I (TCGA) would be expected to detect polymorphisms with a higher efficiency than those that do not contain CpG in their recognition sequences. In fact, all clones that we obtained detected polymorphisms in either *Msp*I or *Taq*I digested DNA, except for OS-7 that detected the insertion/deletion polymorphisms.

As for clones obtained from pUC 18 library, we tested 45 DNA clones, each carrying a single copy DNA fragment; we obtained seven polymorphic DNA clones including the two which were not listed in Table 1, because of low frequency of polymorphisms. If more enzymes had been employed, more clones could have detected polymorphisms. However, in spite of using only four enzymes, this efficiency of detecting polymorphism was consistent with those reports by others (Barker *et al.*, 1984; Cavenee *et al.*, 1984).

Although more than 300 polymorphic DNA markers have been reported (Willard *et al.*, 1985), most of them have only two alleles. As it is important for linkage study of genetic diseases that the more proband or affected parent is heterozygous at a polymorphic locus, probes having many alleles are valuable markers.

Differences in the length of a particular restriction fragment could result from single base-pair changes. In addition to this mechanism, insertion or deletion of short repetitive blocks of DNA could alter its size as has been documented for the loci D14S1 (Wyman and White, 1980), INS (Ullrich *et al.*, 1980; Bell *et al.*, 1981), and HRAS (Goldfarb *et al.*, 1982).

Cloned polymorphic DNA fragments can be mapped to specific chromosomes using human-mouse somatic cell hybrids (de Martinville *et al.*, 1983; Balazs *et al.*, 1984; Naylor *et al.*, 1984) and regional localization can be determined by *in situ* hybridization (Harper and Saunders, 1981) or using human somatic cells having chromosomal aberrations (Cavenee *et al.*, 1984). Now we are establishing the chromosomal localization of nine cloned polymorphic DNA fragments. We determined that one out of nine, OS-4, was located on chromosome 18 (Tateishi *et al.*, 1986). It will not be long before all these clones can be mapped to specific chromosomes.

Acknowledgment We are grateful to Dr. T. Maniatis (Harvard University) for kindly providing a human DNA library. We thank Y. Akiyama for her excellent assistance in preparing the manuscript. This investigation was supported in part by a Grant-in-Aid for Cancer Research, and for Special Project Research, Cancer-Bioscience, from the Ministry of Education, Science and Culture of Japan and by Osaka Kidney Foundation (OKF85-D11).

REFERENCES

Balazs, I., Purrello, M., Alhadeff, B., Grzeschik, K.-H., and Szabo, P. 1984. Isolation and subregional mapping of a human cDNA clone detecting a common RFLP on chromosome 12. *Hum. Genet.* **68**: 57–61.

Barker, D., Schäfer, M., and White, R. 1984. Restriction sites containing CpG show a higher

frequency of polymorphism in human DNA. *Cell* **36**: 131–138.

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

Botstein, D., White, R.L. Skolnick, M., and Davis, R.W. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* **32**: 314–331.

Cavenee, W., Leach, R., Mohandas, T., Pearson, P., and White, R. 1984. Isolation and regional localization of DNA segments revealing polymorphic loci from human chromosome 13. *Am. J. Hum. Genet.* **36**: 10–24.

Cooper D.N. 1983. Eukaryotic DNA methylation. *Hum. Genet.* **64**: 315–333.

Cooper, D.N., Smith, B.A., Cooke, H.J., Niemann, S., and Schmidtke, J. 1985. An estimate of unique DNA sequence heterozygosity in the human genome. *Hum. Genet.* **69**: 201–205.

de Martinville, B., Schäfer, M., White, R., and Francke, V, 1983. Chromosomal assignment of three random RFLP loci defined by base-pair changes in *Msp*I sites. *Mol. Biol. Med.* **1**: 415–424.

Feder, J., Yen, L., Wijsman, E., Wang, L., Wilkins, L., Schroder, J., Spurr, N., Cann, H., Blumenberg, M., and Cavalli-Sforza, L.L. 1985. A systematic approach for detecting high-frequency restriction fragment length polymorphism using large genomic probes. *Am. J. Hum. Genet.* **37**: 635–649.

Goldfarb, M., Shimizu, S., Perucho, M., and Wigler, M. 1982. Isolation and preliminary characterization of a human transforming gene from T24 bladder carcinoma cells. *Nature* **296**: 404–409.

Gusella, J.F., Wexler, N.S., Conneally, P.M., Naylor, S.L., Anderson, M.A., Tauzi, R.E., Watkins, P.C., Ottina, K., Wallace, M.R., Sakaguchi, A.Y., Young, A.B., Shoulson, I., Bonilla, E., and Martin, J.B. 1983. A polymorphic DNA marker genetically linked to Huntington's disease. *Nature* **306**: 234–238.

Hanahan, D. and Meselson, M. 1980. Plasmid screening at high colony density. *Gene* **10**: 63–67.

Harper, M.E. and Saunders, G.F. 1981. Localization of single copy DNA sequences on G-banded human chromosomes by *in situ* hybridization. *Chromosoma* **83**: 431–439.

Helentjaris, T. and Gesteland, R. 1983. Evaluation of random cDNA clones as probes for human restriction fragment length polymorphisms. *J. Mol. Appl. Genet.* **2**: 237–247.

Knowlton, R.G., Cohen-Haguenauer, O., Van Cong, N., Frézal, J., Brown, V.A., Barker, D., Braman, J.C., Schumm, J.W., Tsui, L.-C., Buchwald, M., and Donis-Keller, H. 1985. A polymorphic DNA marker linked to cystic fibrosis is located on chromosome 7. *Nature* **318**: 380–382.

Litt, M. and White, R.L. 1985. A highly polymorphic locus in human DNA revealed by cosmid-derived probes. *Proc. Natl. Acad. Sci. USA* **82**: 6206–6210.

Naylor, S.L., Sakaguchi, A.Y., Barker, D., White, R., and Shows, T.B. 1984. DNA polymorphic loci mapped to human chromosomes 3, 5, 9, 11, 17, 18, and 22. *Proc. Natl. Acad. Sci. USA* **81**: 2447–2451.

Reeders, S.T., Breuning, M.H., Davis, K.E., Nicholls, R.D., Jarman, A.P., Higgs, D.R., Pearson, P.L., and Weatherall, D.J. 1985. A highly polymorphic DNA marker linked to adult polycystic kidney disease on chromosome 16. *Nature* **317**: 542–544

Southern, E.M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98**: 503–517

Tateishi, H., Nishisho, I., Miki, T., Takai, S., and Honjo, T. 1986. An anonymous single copy chromosome 18 probe associated with a frequent RFLP. *Nucl. Acids Res.* **14**: 1926.

Ullrich, A., Dull, T.J., Gray, A., Brosius, J., and Sures, I. 1980. Genetic variation in the human insulin gene. *Science* **209**: 612–615.

Van der Ploeg L.H.T. and Flavell, R.A. 1980. DNA methylation in the human γ,δ,β-globin locus

in erythroid and nonerythroid tissues. *Cell* **19**: 947–958.

Vogel, F. 1972. Non-randomness of base replacement in point mutation. *J. Mol. Evol.* **1**: 334–367.

Vogel, F. and Kopun, M. 1977. Higher frequencies of transitions among point mutations. *J. Mol. Evol.* **9**: 159–180.

Wainwright, B.J., Scambler, P.J., Schmidtke, J., Watson, E.A. Law, H.-Y., Farrall, M., Cook, H.J., Eiberg, H., and Williamson, R. 1985. Localization of cyctic fibrosis locus to human chromosome 7 cen-q22. *Nature* **318**: 348–385.

White, R., Woodward, S., Leppert, M., O'Connell, P., Hoff, M., Herbst, J., Lalouel, J.-M., Dean, M., and Woude, G.V. 1985. A ceosely linked genetic marker for cystic fibrosis. *Nature* **318**: 382–384.

Willard, H.F., Skolnick, M.H., Pearson, P.L., and Mandel, J.-L. 1985. *Cytogenet. Cell. Genet.* **40**: 360–489.

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