IDENTIFICATION OF DNA POLYMORPHISMS

by

PAUL CURTIS WATKINS

B.S., Michigan State University
(1974)

SUBMITTED TO THE DEPARTMENT OF
INTERDISCIPLINARY SCIENCE
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE
DEGREE OF

MASTER OF SCIENCE

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 1982

Signature of Author_____
Department of Interdisciplinary Science
January 11, 1982

Certified by_____
David E. Housman
Thesis Supervisor

Accepted by_____
John Buchanan
Chairman, Departmental Graduate Committee

IDENTIFICATION OF DNA POLYMORPHISMS

by

Paul C. Watkins

Submitted to the Department of Interdisciplinary Science
on January 11, 1982 in partial fulfillment of the requirements
for the degree of Master of Science

ABSTRACT

Three approaches were used to identify DNA polymorphisms in humans by examining the DNAs of a number of individuals within Huntington's Disease pedigrees as well as random individuals for restriction fragment length polymorphisms (RFLPs). The first approach used randomly chosen, cloned single-copy DNA segments from four types of human recombinant DNA libraries as hybridization probes to detect by Southern blot analysis, RFLPs generated by digestion of total human DNA with Eco RI, Sac I, or Hind III. 14 hybridization probes were used to analyze the DNA from panels of 6 or 8 individuals. 133 enzyme sites per individual were examined accounting for 798 base-pairs (bp) of DNA sequence. A total of 3 variant restriction enzyme patterns were found with 3 different hybridization probes: 2 variant patterns result from differences in Hind III cleavage sites, one variant pattern results from the elimination of an Eco RI cleavage site. This Eco RI variant is polymorphic with at least two alleles present and an allele frequency for the absence of an Eco RI site of 0.56. The second approach used restriction enzymes as probes for DNA polymorphisms by examining cloned homologous DNA sequences isolated from human genome libraries made from two individuals. DNA segments of 13-17 kilobase-pairs (kb) of Eco RI-digested human genomic DNA from two individuals were inserted into the bacteriophage lambda vector CH4A*. One recombinant phage clone from each library was randomly chosen. A single-copy DNA subfragment was isolated and used as a hybridization probe to isolate from the other library the homologous cloned DNA sequence (matching clone). Two sets of matching clones were isolated and analyzed with 22 restriction enzymes for restriction fragment length differences. One set of matching clones is identical in all 86 restriction enzyme sites examined, accounting for a total of 407 bp of DNA sequence. A total of 99 restriction enzyme sites (459 bp) in the other set of matching clones was analyzed. Four variant restriction enzyme patterns were found: One with Alu I and Taq I, and two with Hae III. The Taq I variant pattern results from the elimination of a Taq I cleavage site. This Taq I cleavage site variant is polymorphic with at least two alleles present with an allele frequency for the absence of a Taq I site of 0.7. The third approach used cloned middle repetitive DNA probes to detect variant restriction fragment length patterns, by Southern blot analysis, in specific size classes of DNA from two individuals. Two repetitive probes were used to examine patterns generated by 3 different

combinations of restriction enzymes. 7 variant patterns were observed for an estimated 6400 bp of DNA examined. All three approaches appear to be useful for identifying DNA polymorphisms. The frequencies of restriction fragment length variants can be used to estimate the amount of genetic variation in man. The identification of restriction fragment length polymorphisms should contribute towards the construction of a genetic linkage map of the human genome and in resolving modes of inheritance of genetic disease loci.

Thesis Supervisor: Dr. David Housman

Title: Associate Professor of Biology

ACKNOWLEDGMENTS

I would like to extend my thanks to the many individuals who contributed to the successful completion of this thesis. I am particularly indebted to David Housman for his generous support, advice, suggestions, and thoughtful supervision. I would also like to direct special thanks to Jim Gusella, for his expert technical advice on the design of experiments, which strongly influenced the outcome of this work. I am also grateful to Mary Ann Anderson, whose skill with cell culture techniques insured an adequate supply of DNA for experiments; and to Rudy Tanzi, whose considerable practical experience was an invaluable asset. Finally, I would like to extend a very special thanks to my wife Pat, whose constant encouragement, infinite patience, and enthusiastic help, was at all times, most appreciated.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE

CLONED SINGLE-COPY DNA PROBES FOR DNA POLYMORPHISMS

RATIONALE

Although extensive polymorphism in human proteins, representing biochemical variations such as isozymes and cell surface markers, has been demonstrated (1), little information exists on genetic variation at the level of DNA sequence in man. Polymorphism in DNA sequence can now be analyzed directly through the use of recombinant DNA techniques. The ability to identify DNA polymorphisms by using sequence-specific restriction endonucleases obviates the requirement for polymorphisms to be expressed as biochemical phenotypes in order to be detected and used as genetic markers in linkage studies. Restriction fragment length polymorphisms (RFLPs) can result from single base changes in the DNA sequence of a restriction endonuclease recognition site, thereby eliminating or creating a cleavage site. DNA rearrangements such as insertions or deletions can also alter restriction fragment length. Several studies have demonstrated RFLPs at arbitrary loci in the human genome (2,3), on the short arm of chromosome 11 in the region coding for the β-related globin polypeptides (4-7), in the α globin gene cluster on chromosome 16 (8,9), in the gene coding for the α subunit of the family of human glycoprotein hormones (10), in the 5'-flanking region of the human insulin gene (11), in the variable region of human immunoglobin heavy chains (12), and the constant region of λ light chains (13). RFLPs have also been discovered in the mitochondrial DNA of humans

(14,15), mammals (16-19), and apes (20), where they have been used to show the maternal inheritance of mitochondrial DNA and to estimate the level of intraspecific DNA sequence variation.

Polymorphic marker loci, identified as variants in the lengths of DNA fragments produced by restriction enzymes, should be useful for genetic linkage studies for the following reasons:

(i) A large number of arbitrary genetic sites should be defined by RFLPs and the isolation of specific genes would not be required to generate these markers.

(ii) RFLPs are inherited as simple Mendelian codominant markers (3) with special exceptions that could result from the activity of DNA modifying enzymes.

(iii) Individuals can be easily tested for RFLPs. Although this study used lymphocyte cell lines as the source of human DNA, lymphocyte DNA isolated from peripheral blood should be sufficient.

The focus of this thesis is to evaluate three independent approaches to the isolation and identification of DNA polymorphisms. The three methods are discussed separately in the following three chapters. In Chapter 1, random cloned single-copy DNAs from four different types of human recombinant DNA libraries are used as hybridization probes to detect RFLPs in the DNA of individual humans. An alternative approach to identify RFLPs is used in Chapter 2, where a large number of restriction endonucleases are used to exhaustively probe homologous DNA sequences isolated from the genomic libraries of two individuals. In Chapter 3, two types of middle repetitive DNAs are used as hybridization probes to search for RFLPs by comparing restriction digests of the DNA of two individuals.

The first method,described in Chapter 1, uses radiolabelled cloned recombinant DNA as a hybridization probe to compare the DNA restriction fragment pattern of a number of individuals. RFLPs are revealed as variant restriction fragment lengths in a Southern blot analysis (21) when individuals are screened with the same hybridization probe. The cloned DNAs used as probes are entirely single-copy. Repetitive DNA in the hybridization probe would hybridize to a large number of unlinked DNA sequences and prevent the resolution of the hybridization of single-copy DNA to a single genetic locus.

Four different libraries of human recombinant DNA are used as sources of hybridization probes to screen for RFLPs. Each library has properties relevant to this purpose. Only phage carrying single-copy human DNA are suitable as hybridization probes; this places a major limitation on the selection of clones as probes. A library of human DNA segments of 15-20 kb cloned in the λ phage vector Charon 4A (22) permits the use of long DNA sequences as probes which allows many restriction sites to be examined in a single experiment; however, only 1%-3% of the phage clones contain no repetitive DNA. Clones from a second library of smaller human DNA segments (1.5 kb) inserted in the λ phage vector Charon 16A sample fewer restriction sites, but many more clones carrying exclusively single-copy DNA can be isolated. Both of these libraries should contain DNA segments representative of the entire human genome. A third library of human DNA sequences was constructed by enzymatically copying HeLa cell mRNA into DNA (cDNA) and inserting it into the plasmid vector pBR322 (23). The human DNA inserts of this library are each about 300-400 bp long and reflect those human DNA sequences that are transcribed into mRNA. This library represents a special subset of

human DNA sequences and might be expected to contain DNA sequences that are more highly conserved. Clones from the fourth library of human DNA sequences carry 2 kb human single-copy DNA inserts from specific regions of human chromosome 11. Any RFLPs revealed by hybridization of these probes offer the advantage of a specific chromosomal localization.

The library of human DNA fragments of 15-20 kb inserted in the λ phage vector Charon 4A (22) was screened with a repetitive DNA probe for phage carrying only single-copy human DNA inserts. Four such phage were identified and used as hybridization probes to analyze the restriction endonuclease-digested DNA of 6 individuals. Hybridization probes from the human genomic library of 1.5 kb DNA fragments inserted in Charon 16A, the cDNA library made from HeLa cell mRNA, and phage clones from specific regions of human chromosome 11 were also used to screen the same panel of individuals for RFLPs. A total of three different variant restriction fragment length patterns were detected out of the 133 cleavage sites examined per individual. One of the variants showed a polymorphic variation for the presence/absence of an endonuclease Eco RI site and is inherited in a manner consistent with Mendelian inheritance when restriction fragment length patterns of individuals from within a Huntington's Disease pedigree are examined.

MATERIALS AND METHODS

Recombinant DNA Libraries. Four human recombinant DNA libraries were the sources of single-copy cloned human DNA used as hybridization probes to screen for restriction fragment length polymorphisms (RFLPs). Four

phage clones (designated G5, G6, G9, and G12) containing single-copy human DNA were isolated from a human fetal liver DNA library consisting of 15-20 kb human DNA inserts cloned in the λ bacteriophage vector Charon 4A (22). Three clones containing human single-copy DNA (HS3, HS6, HS16) from a library of 1.5 kb human HeLa cell DNA fragments inserted in the λ phage vector Charon 16A were provided by J. Gusella. Four clones (CS1, ES1, BS2, DS2) containing 2 kb single-copy DNA from human chromosome 11 inserted in Charon 16A were also provided by J. Gusella. Three clones were isolated from a cDNA library made from HeLa cell mRNA cloned in the plasmid vector pBR322 as described by D. Kernitt (23).

Screening Human DNA Library for Single-Copy Clones. The recombinant DNA library containing human DNA segments of 15-20 kb inserted into Charon 4A (22) was plated at dilutions to yield single plaques. Random plaques were picked to a fresh lawn of *E. coli*, strain LE 392 (supplied with the λ phage vector gt wes by P. Leder) and oriented in a grid arrangement. These plates were transferred to nitrocellulose filters for *in situ* plaque hybridization as described by Benton and Davis (24). Dry filters (88 mm, Millipore HAHY) were placed on the lawn of cells and phage were allowed to adsorb for approximately 3 minutes. Filters were removed from the agar and phage were denatured and fixed *in situ* by placing the filters in 0.1N NaOH and 1.5M NaCl for twenty seconds. Filters were then neutralized by placing them in 0.2M Tris-HCl, pH 7.5, and 2X SSCP for 20 seconds (1X SSCP: 120 mM NaCl, 15 mM sodium citrate, 13 mM KH_2PO_4, 1 mM EDTA, pH 7.2). Filters were allowed to air dry and were then baked in a vacuum at 80°C for 1.5 hrs. Filters were hybridized to

nick-translated total HeLa DNA to identify clones containing human repetitive DNA in the following manner: Filters were soaked for 3 hrs at 65°C in 250 ml of 4X SET [1X SET: 30 mM Tris-HCl (pH 8.0), 150 mM Nacl, 2 mM EDTA], 10X Denhardt's solution (0.2% bovine serum albumin, 0.2% polyvinylpyrollidone, 0.2% Ficoll), 50 mM $NaPO_4$, 0.1% SDS. Prehybridization of filters was carried out in a sealed plastic bag containing 10 ml of the same solution plus 100 μg/ml denatured calf thymus DNA for 1.5 hrs at 65°C. Filters were hybridized by adding to the prehybridization solution $0.5\text{-}2.0\times10^6$ cpm ^{32}P-labelled probe [specific activity 2×10^8 cpm/μg DNA prepared by nick translation of HeLa DNA (25)]. Hybridization was for 18-24 hrs at 65°C. Filters were washed at 65°C in the following 500 ml solutions: 3X SET, 0.1% SDS for 1 hr, two successive washes in 2X SET, 0.1% SDS for 1 hr, two successive washes in 1X SET, 0.1% SDS for 1 hr. Filters were air dried, covered with plastic wrapping and placed against XAR-5 x-ray film (Kodak) for autoradiography at -70°C with a Dupont Cronex intensifier screen. X-ray film was exposed for 12-48 hrs. Putative single-copy clones were identified by their lack of hybridization to the HeLa DNA hybridization probe.

Preparation of Recombinant Phage DNA. A rapid "mini" preparative method for isolating small quantities of phage DNA provided a convenient method for screening a large number of clones. Single-copy clones were plaque purified, adsorbed to E. coli LE392, and grown as 50 ml broth cultures overnight at 37°C. A few drops of chloroform was added to the culture, which was then clarified by low speed centrifugation in a Sorvall SS-34 rotor at 7,000 rpm for 15 minutes. To the supernatant was added

approximately 0.1 mg each of DNase I (Sigma) and RNase-A (Sigma). After incubation at 37°C for 5 minutes the phage lysate was placed on ice for 1 hr. Phage were pelleted in a Sorvall SS-34 rotor at 16,000 rpm for 3 hrs. The supernatant was decanted and the phage pellet resuspended in 0.2 ml extraction buffer [100 mM Tris-HCl (pH 7.9), 0.3 M NaCl]. The phage DNA was extracted once with an equal volume of phenol and twice with equal volumes of chloroform-isoamyl alcohol (24:1). Phage DNA was precipitated with two volumes of absolute ethanol overnight at -20°C. Ten to fifteen micrograms of DNA were routinely obtained.

Larger quantities of phage DNA were prepared from 500 ml overnight broth cultures of phage-infected E. coli. To the phage lysate was added 5 ml chloroform and 30 grams of NaCl. The phage lysate was clarified by centrifugation in a Sorvall GSA rotor at 6,000 rpm for 20 minutes. 35 grams of polyethylene glycol was added to the lysate which was incubated at 4°C overnight. The phage precipitate was centrifuged in a Sorvall GSA rotor at 6,500 rpm for 30 minutes. The precipitate was resuspended to a volume of 10 ml in φ80 buffer [10 mM Tris- HCL (pH 8.0), 0.1M NaCl, 10 mM $MgCl_2$]. To the resuspended precipitate was added 7.5 grams CsCl . Phage were banded by two cycles of equilibrium ultra-centrifugation in a Beckman 65 rotor at 45,000 rpm for 24 hrs. Phage bands were collected by side puncture of the gradient with a 20 gauge syringe needle and dialyzed overnight at 4°C against 4 liters 080 buffer. Phage DNA was extracted as for "mini" preparations as described above. 200 to 500 micrograms of DNA were routinely obtained.

Preparation of Recombinant Plasmid DNA. Rapid "mini" preparations of bacterial plasmid DNA were prepared in the following manner: Random

clones from a HeLa cell cDNA library (23) were picked and grown as 50 ml overnight broth cultures in media supplemented with tetracycline, 20 μg/ml. Bacteria were pelleted in a Sorvall SS-34 rotor at 3,000 rpm, 10 minutes. Supernatants were decanted; the pellet resuspended in 1.0 ml STET buffer [8% sucrose, 5% Triton X-100, 50 mM EDTA, 50 mM Tris-HCl (pH 8.0)] and 250 μl of a 10 mg/ml stock solution of a freshly prepared lysozyme (Sigma) added. The solution was rapidly brought to a boil over a flame then maintained in a boiling water bath for 15-30 seconds. The solution was immediately centrifuged at 12,000 x g for 10 minutes at room temperature. The supernatant was collected and nucleic acid was precipitated overnight at -20°C after adding an equal volume of isopropanol. 20 micrograms of plasmid DNA were routinely obtained.

Source and Preparation of Human DNA. Human DNA was isolated from lymphocyte cell lines of 8 individuals belonging to a Huntington's Disease pedigree (Human Genetic Mutant Cell Repository, Camden, New Jersey). Cells were digested in buffer containing SDS and Proteinase K. DNA was isolated by phenol and chloroform extraction and isopropanol precipitation according to standard methods (26). Potential restriction fragment length polymorphisms (RFLPs) were checked against a larger panel which consisted of 19 individuals from within a Huntington's Disease pedigree (Human Genetic Mutant Cell Repository, Camden, New Jersey), two Venezuelans (from a family with a high incidence of Huntington's Disease), a Ghanian (GM2064, Human Genetic Mutant Cell Repository, Camden, New Jersey), and 3 individuals with family histories of neurofibromatosis.

Restriction Enzyme Digestion and Electrophoresis of DNA. Restriction enzymes were obtained from Bethesda Research Laboratories (Rockville, MD), Boehringer Mannheim, and New England Biolabs and used as directed. Human DNA was digested with Eco RI, Sac I, or Hind III at a DNA concentration of 500 μg/ml. Enzyme digestions were carried out for 2 hrs at 37°C. Restriction enzyme was inactivated by incubation for 5 minutes at 65°C. 10 μg aliquots of restriction enzyme-digested DNA were loaded onto a 1% agarose (FMC, Rockland, ME) gel. DNA was electrophoresed at 40 volts for 15-18 hrs. The initial screening for RFLPs consisted of digestion and electrophoresis of DNA from 6 individuals, each digested with Eco RI, Sac I, and Hind III. In order to visualize small differences in mobility that might characterize RFLPs, restriction digests were applied to gels so that DNAs digested with the same enzyme were in adjacent lanes.

Transfer and Hybridization of DNA. The gel was incubated for 45 minutes in 1M NaOH followed by two 20 minute incubations in 1M Tris-HCl (pH7.6), 1.5M NaCl. Human DNA was transferred from gel to nitrocellose by the method of Southern (21). After transfer filters were baked under reduced pressure for 2 hrs at 80°C. Filters were prehybridized for 3 hrs at 65°C in 10 ml of a solution containing 6XSSC (1XSSC: 0.15M NaCl, 0.015M Na Citrate), 10X Denhardt's Solution, 50 mM $NaPO_4$, 0.1% SDS, and 100 μg/ml denatured calf thymus DNA. Filters were rinsed in 500 ml of 3XSSC and air dried. Hybridization to ^{32}P-labelled probe was carried out in a sealed plastic bag in a 4 ml solution containing 6XSSC, 10X Denhardt's, 50 mM $NaPO_4$, 0.1% SDS, 100 μg/ml calf thymus DNA, and $10-15X10^6$ cpm ^{32}P-labeled hybridization probe (S.A. approx. $1.0X10^8$

cpm/μg DNA). Hybridization was at 65°C for 24 hrs. Filters were then washed at 65°C in 500 ml portions of the following: 3XSSC, 0.1% SDS for 30 minutes; two times for 30 minutes with 2XSSC, 0.1% SDS; two times for 30 minutes with 1XSSC, 0.1% SDS. Filters were air dried , wrapped in plastic and placed against XAR-5 (Kodak) x-ray film for autoradiography at -70°C with a Dupont Cronex intensifier screen for 1-4 days.

Preparation of Hybridization Probes. Purified human single-copy DNA inserts, whole recombinant phage DNA, or recombinant plasmid DNA was used as hybridization probes. Human DNA inserts were obtained by digestion of the recombinant phage with the appropiate restriction endonuclease to separate the human DNA insert from λ end fragments. The restriction digests were then fractionated by agarose gel electrophoresis, followed by electroelution of DNA from the agarose. DNA was labelled with ^{32}P to a specific activity of approximately 0.5-2.0 X 10^8 cpm/μg DNA by the procedure of Rigby et al. (25).

RESULTS

Detection of RFLP in Human DNA

Four types of hybridization probes were used to screen restriction endonuclease digests of total human DNA for restriction fragment length differences. Recombinant Charon 4A phage containing 15-20 kb single-copy human DNA inserts bounded by synthetic Eco RI sites were isolated from a library of human DNA segments constructed by Lawn et. al (22).

These clones were selected by their lack of hybridization to ^{32}P-labelled total human DNA probe (HeLa cell DNA, S.A. 1X10^{8} cpm/μg DNA) when nitrocellulose filters of phage plaques were screened using the procedure of Benton & Davis (24). Only phage with completely single-copy human DNA inserts would fail to hybridize under the conditions used. Although most human DNA segments of 15-20 kb contain repetitive DNA sequences, following the general pattern of interspersed unique and repeated elements found in most eukaryotes (27), it has been shown that this library contains phage clones with some human single-copy DNA sequences of greater than 13 kb (3). Four such phage clones were isolated and designated G5, G6, G9, and G12. Three other types of hybridization probes were used: Three phage clones containing human (HeLa cell) single-copy DNA inserts of 1.5 kb inserted into the lambda phage vector Charon 16A; three clones randomly picked from a human (HeLa cell) cDNA library in plasmid pBR322 (23); and four phage clones of 2 kb single-copy DNA from human chromosome 11 inserted into Charon 16A. Whole phage DNA, plasmid DNA, or purified human DNA inserts were labelled with ^{32}P to a specific activity of approximately 1X10^{8} cpm/microgram DNA.

Human DNA from individuals to be examined for restriction fragment length variants was digested to completion with either Eco RI, Sac I, or Hind III. Restriction digests were then fractionated by agarose gel electrophoresis, and transferred to nitrocellulose by the method of Southern (21). A typical experiment is shown in Figure 1 and illustrates the hybridization of G5 to restriction digests of 6

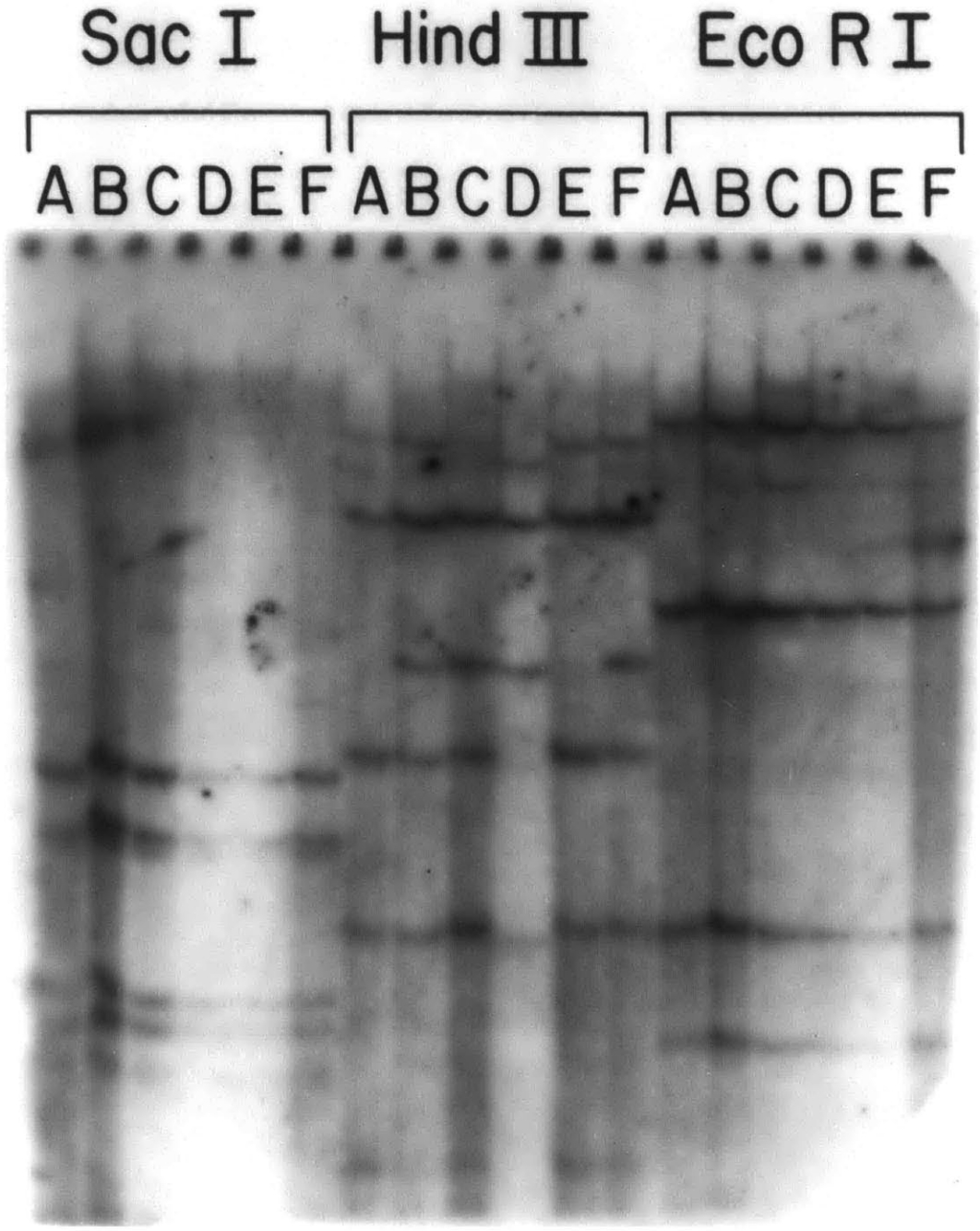

Sac I
Hind III
Eco R I
A B C D E F A B C D E F A B C D E F

Figure 1: Hybridization of G5 Probe to Eco RI, Hind III, and Sac I Digested DNAs.

DNA extracted from the indicated cell line was digested with Eco RI, Hind III, or Sac I, fractionated on a 1 % agarose gel, transferred to nitrocellulose and hybridized to ^{32}P-labelled G5 phage DNA.

Lanes: A, GM498; B, GM491; C, Gus 13; D, Gus 5; E, GM4219; F, GM4207

individuals. The restriction fragment patterns of Sac I-digested DNA or Eco RI-digested DNA for all 6 individuals are the same. Hind III-digested DNA displays a variant restriction pattern with 2 individuals (B&F) lacking an upper DNA fragment present in the other 4 individuals; individual C is missing a DNA fragment in the middle of the autoradiograph that is shared by the other 5 individuals. The uppermost Hind III-digested DNA fragment of individual C also shows a variant restriction fragment length when compared to the other 5 individuals. These preliminary results suggest that a base substitution has probably resulted in the presence/absence of at least one Hind III restriction site. A restriction fragment length variation due to a deletion or insertion of a block of DNA sequence would most likely be reflected in variant patterns displayed by Eco RI-digested DNA and Sac I-digested DNA.

The results of the hybridization of a total of 14 single-copy clones to enzyme-digested human DNA are summarized in Table I. 133 restriction endonuclease cleavage sites per individual were examined accounting for a total of 798 bp of DNA sequence per individual. A total of three variant restriction enzyme patterns were observed. Assuming that the presence or absence of a restriction site is the result of a single base pair change, a total of 3 out of 798 bp screened show polymorphic variation. The frequency of polymorphism estimated from this data is 3/798=0.0038.

Inheritance of Eco RI RFLP Detected by Clone G6

Clone G6 isolated from a library of human DNA segments inserted in

Table I. Hybridization of Random Single Copy-Clones to EcoR1, Sac I, and Hind III-Digested Genomic DNAs.

No. of Enzyme Recognition Sites in Hybridizing Fragments (No. of Individuals Tested)				Total No. Enzyme Sites per Individual	Total No. Base Pairs	No. of Differences
Clone	EcoR1	Sac I	Hind III			
G6	4 or 5 (23)	5(4)	7(12)	17	102	1 EcoR1
G12	5(6)	7(6)	5(6)	17	102	
G9	5(6)	4(6)	4(6)	13	78	
G5	7(8)	6(8)	6 or 7 (8)	20	120	1 Hind III
Insert Size:		15 to 20 kb				
HS3	2(6)	2(6)	3(6)	7	42	
HS6	3(6)	2(6)	3(6)	8	48	
HS16	2(6)	2(6)	2(6)	6	36	
Insert Size:		1.5 kb				
pH1	NT	2(6)	2(6)	4	24	
pH2	2(6)	3(6)	2(6)	7	42	
pH3	2(6)	2(6)	2(6)	6	36	
Insert Size:		300 to 400 bp				
CS1	2(2)	3(2)	2(2)	7	42	
ES1	2(6)	2(6)	3(6)	7	42	
BS2	2(6)	2(6)	2(6)	6	36	
DS2	2(6)	2(6)	2 or 4 (11)	8	48	1 Hind III
Insert Size:		2 kb				

No. of Enzyme Sites examined per Individual: 133

No. of Base Pairs in Enzyme Sites per Individual: 798

Charon 4A contains three single-copy human DNA inserts of 3.4 kb, 4.6 kb, and 6.8 kb (data not shown); the total insert is bounded by synthetic Eco RI recognition sites. A variant restriction fragment length pattern is obtained when G6 is used as a hybridization probe in a Southern blot analysis of Eco RI-digested human DNA. The results of the hybridization for 11 individuals is shown in Figure 2. Two restriction fragment length patterns are apparent. One pattern consists of 3 bands with the lower two migrating as a doublet; the other pattern is lacking the lowermost fragment of the doublet. The results are interpreted to indicate the presence of at least two alleles. Individuals with both lower bands are heterozygous for two fragment lengths (alleles) and individuals with only one of the two lower bands are homozygous for a single allele. The pattern of inheritance of restriction fragment lengths for 24 individuals is shown in Table III. The familial relationships of those individuals tested who are members of a Huntington's Disease pedigree are shown in Figures 3 & 4. 9 individuals are homozygous for one restriction fragment length: 8 individuals have only the upper band of the doublet while 1 individual has only the lower band of the doublet. 15 individuals are heterozygous for both restriction fragment lengths. The restriction fragment lengths observed are consistent with their inheritance as Mendelian alleles through 3 generations. No correlation is observed between the inheritance of either restriction fragment length and the autosomal dominant genetic disease trait that is responsible for Huntington's Disease.

Eight unrelated individuals from the above panel of 24 were used to determine the allele frequencies of the Eco RI RFLP. 5 individuals are heterozygous for both restriction fragment lengths, 2 individuals are

homozygous for the upper band of the doublet, and 1 individual is homozygous for the lower band of the doublet. This distribution gives an allele frequency of 0.56 for the upper restriction fragment length.

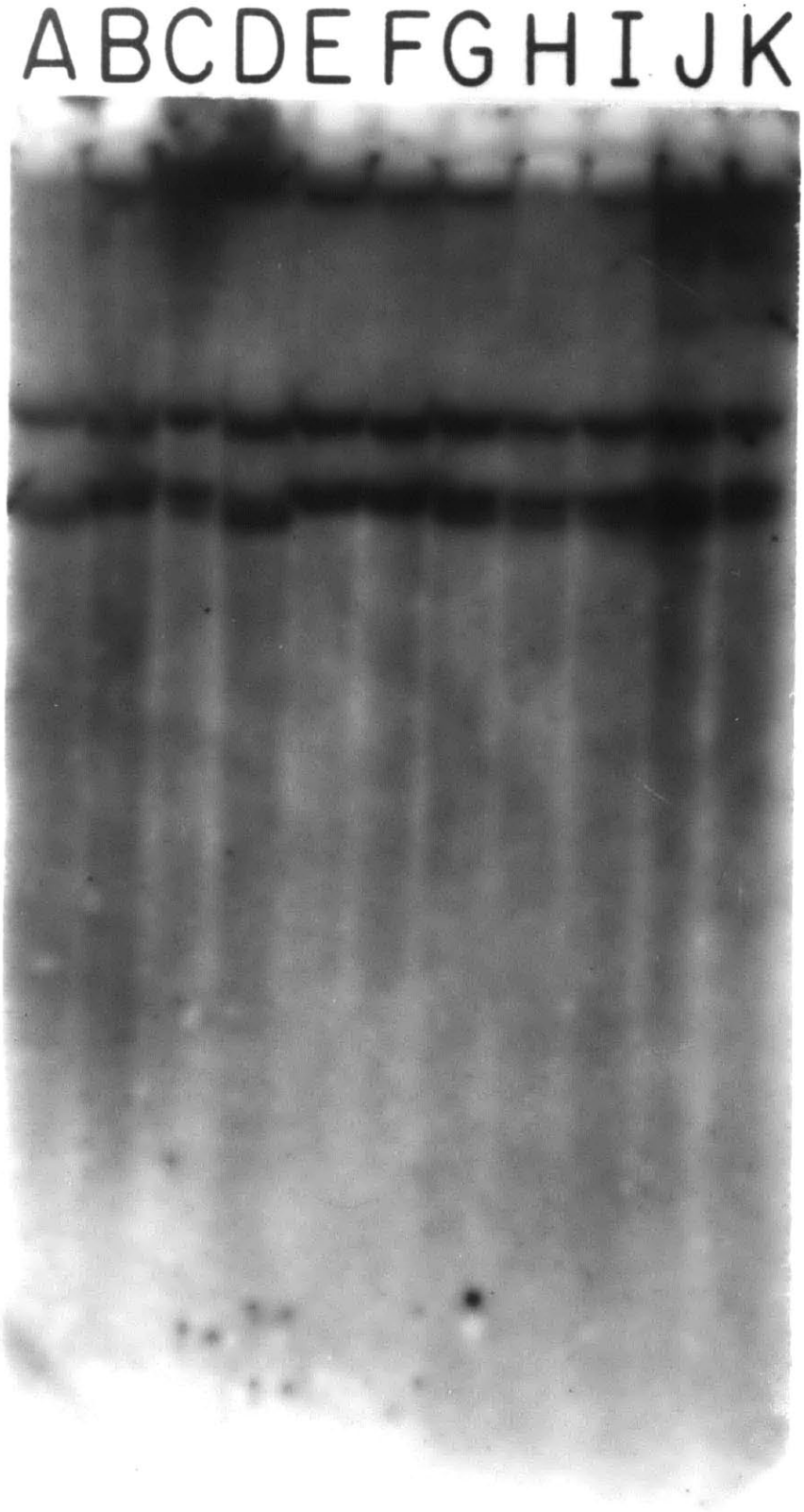

A B C D E F G H I J K

Figure 2: Hybridization of G6 Probe to Eco RI Digested Cell DNAs. DNA extracted from the indicated cell lines was digested with Eco RI, fractionated on a 1% agarose gel, transferred to nitrocellulose and hybridized to ^{32}P-labelled G6 phage DNA.

Lanes: A, GM4207; B, GM4219; C, GM4189; D, GM4191; E, GM4227; F, GM4233; G, GM4197; H, GM4201; I, GM4195; J, GM4223; K, GM4199.

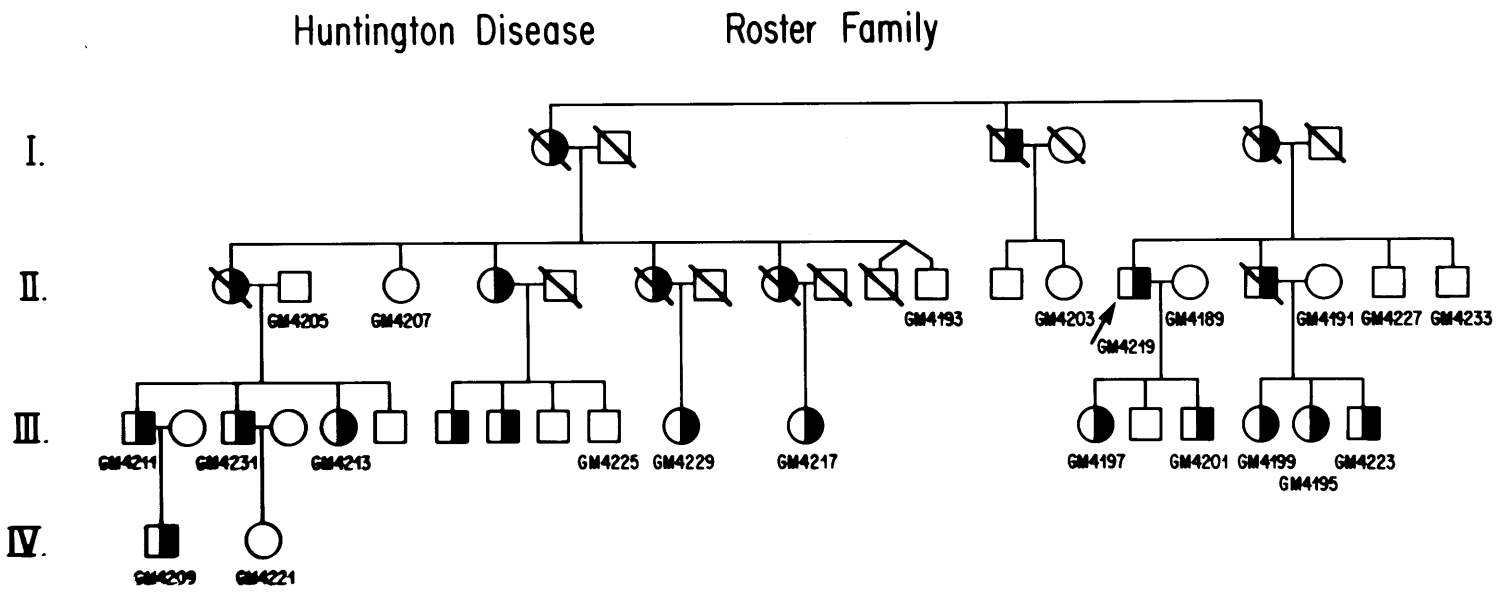

Huntington Disease Roster Family
I.
II.
III.
IV.
GM4205
GM4207
GM4193
GM4203
GM4219
GM4189
GM4191
GM4227
GM4233
GM4211
GM4231
GM4213
GM4225
GM4229
GM4217
GM4197
GM4201
GM4199
GM4223
GM4195
GM4209
GM4221

Figure 3: Diagram Showing Huntington's Disease Pedigree. Circles (females) or squares (males) with numbers represent individuals from whom lymphocyte cell lines were established. Half-filled circles or half-filled squares represent individuals with Huntington's Disease. Slashes designate deceased individuals.

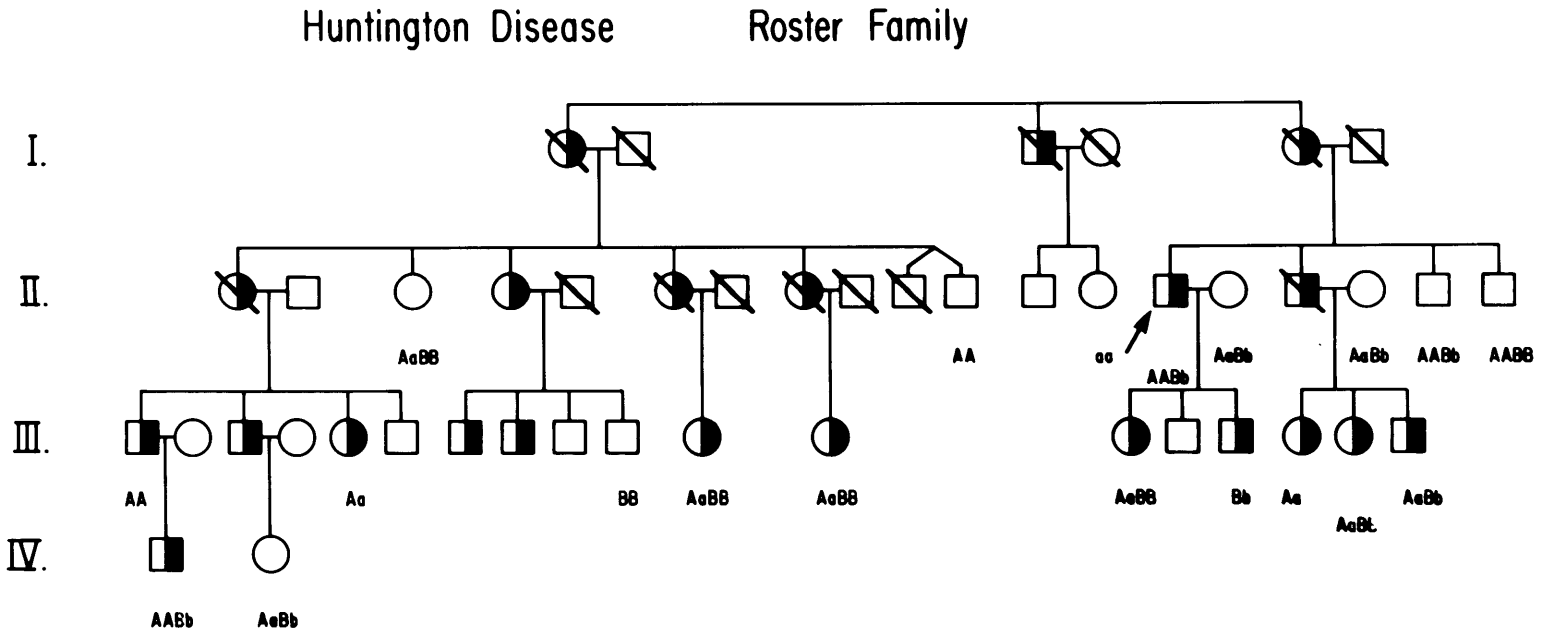

Huntington Disease
Roster Family
I.
II.
III.
IV.
AaBB
AA
aa
AABb
AaBb
AaBb
AABb
AABB
AA
Aa
BB
AaBB
AaBB
AaBB
Bb
Aa
AaBb
AaBb
AABb
AaBb

Figure 4: Diagram Showing Huntington's Disease Pedigree and Inheritance of DNA Polymorphisms.

Circles (females) or squares (males) with numbers represent individuals from whom lymphocyte cell lines were established. Half-filled circles or half-filled squares represent individuals with Huntington's Disease. Slashes designate deceased individuals. A, indicates inheritance of upper DNA segment of Eco RI RFLP detected by G6 probe. a, indicates inheritance of lower DNA segment of Eco RI RFLP detected by G6 probe. B, indicates upper DNA segment (3kb) of Taq I RFLP detected by HDIA-H4 probe. b, indicates lower two DNA segments (2kb & 1kb) of Taq I RFLP detected by HD1A-H4 probe.

CHAPTER 2

IDENTIFICATION OF DNA POLYMORPHISMS IN MATCHING CLONES

RATIONALE

The object of this study was to search for DNA polymorphisms by using restriction endonucleases as probes for DNA sequence differences between homologous DNA sequences isolated from two individuals. Recombinant DNA libraries of human DNA inserted in a λ bacteriophage vector were made from two individuals and served as the source of homologous DNA sequences. Two considerations governed library construction: First, that the human DNA inserts were of sufficient length to allow the assay of many restriction sites with a given enzyme. Second, that the libraries were constructed in a manner that facilitated the screening for homologous DNA sequences. This was accomplished by reducing the complexity of the DNA content of the total human genome by selecting a specific size class (13-17 kb) of restriction enzyme-digested DNA to be inserted into the λ phage vector. The same small percentage of the genome would be represented by the recombinant phage comprising each of the libraries. Single-copy subfragments of randomly isolated clones from each library were used as hybridization probes to find the homologous DNA sequences (matching clones) in the other library.

Two sets of matching clones were isolated and analyzed with 22 restriction endonucleases for RFLPs. One set of clones showed no variation in all 86 cleavage sites (407 bp DNA) examined. The other set of clones displayed variant restriction fragment length patterns when

digested with Alu I, Hae III, and Taq I. The Taq I variant showed a polymorphic variation for the presence/absence of a endonuclease Taq I cleavage site. The Taq I restriction fragment lengths are inherited as Mendelian alleles through three generations of a Huntington's Disease pedigree.

MATERIALS AND METHODS

Construction of Human DNA Libraries HDA & HDB

Enzyme Digestion, Electrophoresis, and Electroelution of Human DNA. DNA from two transformed lymphocyte cell lines (GM4207 & GM4219) was isolated as described in Chapter 1. GM4207 and GM4219 are transformed lymphocyte cell lines obtained from two individuals from within a Huntington's Disease pedigree. Approximately 350 μg of DNA were digested to completion with 450 units of EcoR1 (Boehringer Mannheim) at 37°C for 3 hrs. The enzyme was inactivated at 65°C for 5 minutes. The samples were extracted once with an equal volume of phenol-chloroform-isoamyl alcohol (25:24:1) and once with chloroform-isoamyl alcohol (24:1). DNA was precipitated with two volumes of absolute ethanol at -18°C for 24 hrs. DNA was pelleted by low speed centrifugation. The pellet was dried under reduced pressure then resuspended in 0.2 ml TNE [10 mM Tris-HCl (pH 7.5), 0.2M NaCl, 1 mM EDTA]. Both DNA samples were applied to separate halves of a 1% agarose gel and electrophoresed at 40 volts for 24 hrs. DNA was visualized by staining the gel with ethidium bromide. The gel was sliced horizontally in 6mm slices beginning with the first visible fraction of high molecular weight DNA. Gel slices

were individually placed in dialysis tubing (Spectrum, 23 mm) with a small amount of TE [10 mM Tris-HCl (pH 8.0), 1 mM EDTA]. DNA was electroeluted from the agarose in TE buffer with a current of 75 mA. Electroeluted DNA was passed through a small plug of glass wool in a Pasteur pipet and then concentrated to 0.4 ml by repeated extractions with isobutyl alcohol. DNA samples were dialyzed overnight in collodion bags (Schleicher and Schuell) against 3 liters of TNE. The samples were extracted with phenol and chloroform-isoamyl alcohol as described above. DNA was precipitated with two volumes of absolute ethanol, dried, and resuspended in 30 µl TE.

Preparation of Bacteriophage Lambda DNA. The λ phage vector Charon 4A* (CH4A*) was provided by J. Gusella. The CH4A* vector is a derivative of the EK-2 certified vector Charon 4A (28). The CH4A* vector contains an additional *E. coli lac* 5 substitution (6.9 kb) in place of the *bio* 256 substitution (7.8 kb) found in Charon 4A. End fragments of CH4A* were prepared by digesting 100 µg of phage DNA with 200 units of Eco RI (Boehringer Mannheim) in a volume of 0.5 ml. The DNA was extracted once with phenol and then with chloroform. 100 µg of Eco RI-cleaved DNA was layered on a 5-25% linear NaCl gradient [10 mM Tris-HCl (pH 8.0), 1 mM EDTA] in a Beckman SW41 tube. The gradient was centrifuged at 39,000 rpm for 5 hrs at 20°C. 0.5 ml fractions were collected and analyzed on an agarose gel. Fractions containing CH4A* end fragments were pooled and diluted 5 fold with TE. DNA was precipitated with two volumes of ethanol and pelleted in a Sorvall SW 27 rotor at 20,000 rpm for 1 hr at 4°C. The DNA pellet was dried and resuspended to 0.5 ml with TNE. DNA was transferred to an Eppendorf centrifuge tube, precipitated again with

ethanol, dried, and resuspended to 100 μl with TE.

Ligation and In Vitro Packaging of Recombinant DNA. Eco RI-digested human DNA from GM4207 and GM4219 was electroeluted as described above from one set of parallel gel slices containing DNA in the molecular size range of 13 kb to 17 kb. 0.1 μg of Eco RI-digested human DNA was added to 0.5 μg of purified Eco RI-digested CH4A* end fragments in 25 μl of ligase buffer [100 mM Tris-HCl (pH 7.5), 50 mM NaCl, 5 mM $MgCl_2$, 1mM ATP, 1 mM DTT, 2 μg BSA, 400 units T4 ligase (New England Biolabs)]. The ligation mixtures were incubated at 15°C for 15 hrs. An aliquot was heated to 65°C for 5 minutes and electrophoresed on a 0.8% agarose gel with Hind III-digested lambda cI857Sam7 DNA as a molecular weight standard. Successful ligation was evidenced by the absence of Charon 4A* end fragments (11 and 20 kb) and the presence of concatemeric DNA molecules larger than intact CH4A* DNA.

In vitro packaging of recombinant DNA molecules into phage particles was as described by F. R. Blattner in the detailed protocol that accompanies the Charon lambda phages. This protocol is an adaptation of the procedure of Becker and Gold (29).

Library Amplification. Test packaging of a 4 μl aliquot of the ligation mixtures allowed determination of the in vitro packaging efficiency. The rest of the ligation mixtures were packaged and plated onto a fresh lawn of E. coli at a density of 10,000 phage per 15 cm plate (30 plates total). The top agar layer containing phage plaques was scraped into sterile 50 ml centrifuge tubes. One ml chloroform and 0.1 volumes of 10X φ80 buffer was added to the agar mixture which was then incubated at

4°C for 24 hrs. Agar was removed by centrifugation in an IEC DPR-6000 centrifuge at 7,000 rpm for 30 minutes. Libraries were stored over 0.2 ml chloroform at 4°C.

Hybridization of Libraries to Total Human DNA. An estimation of the percentage of each library consisting of recombinant phage containing human repetitive DNA was obtained by hybridization to total human DNA (HeLa cell DNA). Libraries were plated on E. coli at dilutions to yield approximately 1,000 phage per 15 cm plate. Filters were made according to the procedure of Benton and Davis (24) as described in Chapter 1. $5X10^6$ cpm of ^{32}P-labelled HeLa DNA were hybridized to separate filters under conditions described in Chapter 1. Filters were washed to 1X SET, air dried, wrapped in plastic film and placed against XAR-5 x-ray film (Kodak) and a Dupont Cronex intensifier screen for autoradiography at -70° for 18 hrs.

Screening Libraries for Matching Clones

Sets of matching clones (recombinant CH4A* phage carrying homologous human DNA sequence) were selected from both libraries by hybridization to single-copy DNA fragments isolated from randomly selected clones. Libraries were plated at dilutions to yield approximately 100 pfu per 15 cm plate. 100 μg/ml of top agar of the chromogenic substrate 5-chloro-4-bromo-3-indolyl-β-D-galactoside (XG) was included in the plating medium in order to distinguish recombinant (colorless) from intact CH4A* (blue) phage plaques. Random plaques were picked and plaque purified through a second cycle of adsorption and plating on E. coli. Well

isolated plaques were picked with a sterile Pasteur pipette and phage were eluted into 0.2 ml of adsorption buffer (10 mM $MgCl_2$, 10 mM $CaCl_2$) at room temperature for 3 hrs. Phage were adsorbed to 0.2 ml of a late log phase culture of E. coli at 37°C for 40 minutes, then added to 50 ml of fresh medium and incubated overnight. DNA "mini" preparations were made from phage lysates as described in Chapter 1.

One microgram of phage DNA was digested to completion with Eco RI and each of the following restriction endonuclease combinations: Eco RI plus Bam H1, Eco RI plus Sac I, Eco RI plus Hind III, Eco RI plus Kpn I, Eco RI plus Bgl II, Eco RI plus Pvu II. Enzyme was inactivated at 65°C for 5 minutes and samples were applied to a 0.8% agarose gel for electrophoresis. The gel was transferred to nitrocellulose by the method of Southern (21) and hybridized to $5X10^6$ cpm of ^{32}P-labelled nick-translated HeLa DNA. Transfer and hybridization was as described in Chapter 1. The filter was washed to 1X SSC and subjected to autoradiography as previously described. Single-copy human DNA fragments were identified by their lack of hybridization to ^{32}P-labelled HeLa DNA. Clones for which suitable single-copy DNA fragments had been identified were chosen for large scale DNA preparation by methods described in Chapter 1. 20 μg of phage DNA were digested with the appropriate enzyme combinations and electrophoresed on 1% agarose gels. DNA was visualized by ethidium bromide staining and human single-copy DNA fragments identified. DNA fragments were excised from the gel with a scalpel blade, electroeluted, concentrated with isobutyl alcohol, extracted with phenol and chloroform-isoamyl alcohol, and precipitated with absolute ethanol according to methods specified in Chapter 1.

Approximately 0.5 μg of human single-copy DNA isolated from

preparative agarose gels was labelled with ^{32}P by nick translation by the method of Rigby (25). Human recombinant DNA libraries (13-17 kb Eco RI human DNA fragments cloned in CH4A* as described above) were each plated at densities of 2,000 phage per 15 cm plate (total 10 plates) and filters were made by the method of Benton and Davis (24). $1.5X10^5$ cpm of ^{32}P-labelled human single-copy DNA was hybridized to filters using conditions described in Chapter 1 but included 50 µl HeLa DNA and 20 µl CH4A* DNA as carrier DNA per 10 ml prehybridization and hybridization solutions. Filters were washed to 0.3XSET and subjected to autoradiography. Autorads were examined for areas of intense hybridization and aligned with filters and plates for identification of hybridizing clones. Clones were picked with Pasteur pipettes, adsorbed to E. coli, and plated at dilutions to yield well isolated plaques. Benton-Davis filters were made and hybridized to ^{32}P-labelled human single-copy DNA probe under conditions described above. Those clones showing increased numbers of positive signals after rescreening and autoradiography were selected for comparative restriction endonuclease analysis.

Comparative Restriction Endonuclease Analysis of Matching Clones

DNA from sets of matching clones was prepared by large scale methods described in Chapter 1. DNA was digested with 22 different restriction endonucleases (see Table 2) and patterns were analyzed by electrophoresis in adjacent lanes of agarose gels. Restriction endonucleases that recognize a 6 base pair DNA sequence and cut λ DNA infrequently were used in combination with Eco RI to digest recombinant

phage DNA. Products were analyzed on 0.8%, 1%, or 1.4% agarose gels. A restriction map of CH4A* allowed differentiation of phage DNA fragments from human DNA fragments. Prior to digestion with restriction endonucleases that recognize a 4 or 5 base pair sequence, human DNA inserts were first purified from λ phage DNA. 60 μg of phage DNA was digested with 150 units of Eco RI. Electrophoresis and electroelution of human DNA inserts was according to methods described above. Products were analyzed on 2% agarose gels.

Restriction Endonuclease Map of Clone HD1A

The restriction endonuclease recognition sites for Eco RI, Bam H1, Hind III, Sac 1, and Kpn 1 in clone HD1A were determined by digesting 2 μg of phage DNA separately with each enzyme and with combinations of the above enzymes. Known molecular weights of DNA fragments generated by the above enzymes with end fragments of CH4A* and a Hind III digest of λcI857Sam7 served as molecular weight standards. DNA was visualized by ethidium bromide staining after electrophoresis on a 1% agarose gel.

Hybridization of Subfragments of Clone HD1A to Matching Clones HD1A and HD1B

One microgram of purified human DNA insert from matching clones HD1A and HD1B was digested to completion with Hae III and electrophoresed in adjacent lanes of a 2% agarose gel. DNA was visualized by staining with ethidium bromide and transferred to nitrocellulose by the method of Southern (21) as described in Chapter 1. The filter was hybridized to

subfragments of clone HD1A generated by Hind III plus Eco RI digestion of 30 μg HD1A DNA. DNA was electrophoresed on a 1% agarose gel and human DNA fragments of 6.1 kb, 3.7 kb, 2.0 kb, and 1.9 kb were excised and electroeluted from the agarose. The filter was sequentially hybridized to $2X10^6$ cpm of the 6.1 kb, 3.7 kb, and 2.0 kb ^{32}P-labelled subfragment probes and was washed to 0.01X SSC before each hybridization. Hybridization conditions were essentially as described in Chapter 1, except that the filter was placed in a sealed plastic bag with 10 ml prehybridization solution at 65°C for 1 hr followed by addition of the hybridization probe to the bag. The filter was washed to 0.3XSSC plus 0.1%SDS, air dried, and subjected to autoradiography as previously described.

One microgram of the electroeluted 1.9 kb fragment generated by Eco RI plus Hind III digestion of clone HD1A was digested to completion with Hae III. DNA was electrophoresed on a 2% agarose gel with Hind III digested λcI857Sam7 as a molecular weight standard. A 790 base pair DNA fragment was excised from the gel, electroeluted, and purified according to methods specified in Chapter 1. The DNA fragment was labelled with ^{32}P by nick translation (25) and $1X10^6$ cpm hybridized to Hae III-digested HD1A and HD1B under conditions described above.

Subcloning of HD1A-H4 in Charon 21A

0.25 μg of Hind III-digested Charon 21A and 0.1 μg of the 3.7 kb DNA fragment generated by Hind III digestion of clone HD1A were ligated by using 400 units T4 ligase (New England Biolabs) under conditions described above. The ligation products were packaged *in vitro* (28) and

phage plated on E. coli at dilutions to yield approximately 100 phage per 15 cm plate. Benton-Davis filters were made and hybridized to $1X10^6$ cpm of ^{32}P-labelled 3.7 kb DNA fragment purified from a Hind III digest of clone HD1A. Hybridization conditions were as described above for Benton-Davis filters except that 20 µl of Charon 21A DNA were included in the hybridization mixture. Filters were washed to 1XSET and subjected to autoradiography. A well-isolated and intensely hybridizing phage plaque was selected, adsorbed to E. coli, and grown overnight in a 50 ml culture. DNA was isolated according to a "mini" preparative technique discussed in Chapter 1. 2 micrograms phage DNA was digested with Hind III and electrophoresed on a 1% agarose gel. Successful subcloning was confirmed by observation of a 3.7 kb DNA fragment not observed with Hind III-digested Charon 21A. The recombinant phage that contained the 3.7 kb insert corresponding to the 3.7 kb Hind III digestion product of clone HD1A was designated HD1A-H4.

Hybridization of Subfragments of Clone HD1A to DNA of Individuals

Human DNA was isolated from lymphocyte cell lines from individuals according to methods specified in Chapter 1. To screen individuals for Hae III RFLPs, aliquots of 10 µg of DNA from each of 10 individuals from within a Huntington's Disease pedigree (Human Genetic Mutant Cell Repository, Camden, New Jersey) were digested to completion with Hae III and electrophoresed on a 2% agarose gel. The Taq I RFLP was screened against a panel of 22 individuals, 15 from within a Huntington's Disease pedigree (Camden, New Jersey), 4 individuals with family histories of neurofibromatosis, 2 Venezuelans, and one Ghanian (GM2064, Human Genetic

Mutant Cell Repository, Camden, New Jersey). 10 μg of DNA from each individual were digested to completion with Taq I at 65°C and electrophoresed on a 1% agarose gel. DNA was transferred to nitrocellulose by the method of Southern (21) and hybridized to ^{32}P-labelled subfragments of clone HD1A or to clone HD1A-H4. DNA fragments for hybridization probes were obtained from Hind III-digested HD1A or HD1A-H4 DNA. The 3.7 kb human single-copy DNA insert was electroeluted and purified according to methods specified above. Hybridization conditions were essentially as described in Chapter 1. $13X10^6$ cpm of ^{32}P-labelled 3.7 kb probe were hybridized to filters containing Taq I-digested human genomic DNA. $10X10^6$ cpm of ^{32}P-labelled 3.7 kb probe were hybridized to a filter containing Hae III-digested human genomic DNA. Filters were washed to 0.3XSSC plus 0.1% SDS and subjected to autoradiography.

RESULTS

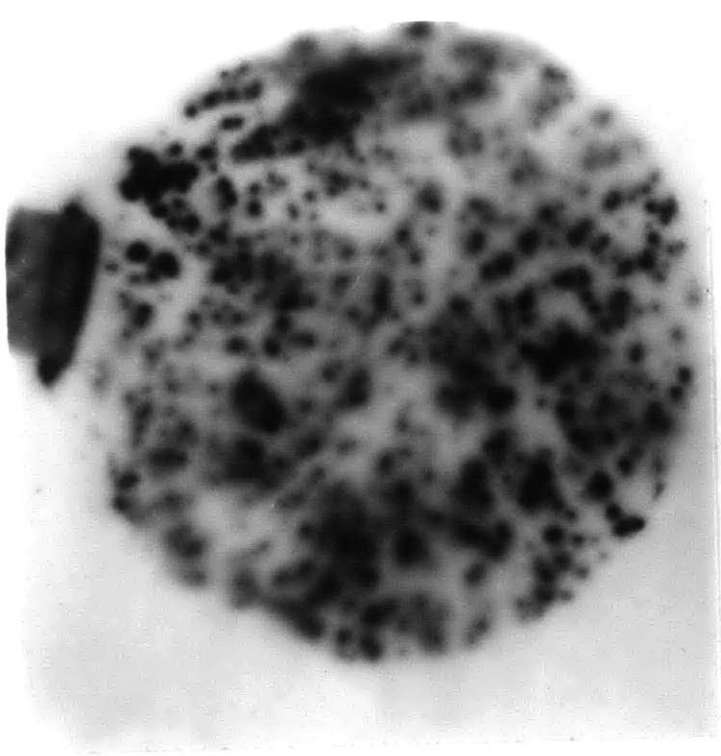
Characterization of Libraries HDA and HDB

Lymphocyte cell lines GM4207 and GM4219 isolated from two individuals within a Huntington's Disease pedigree were used as the source of DNA for the libraries. A specific size class of Eco RI-digested human DNA fragments (13kb-17kb) from the two cell lines was isolated after fractionation of restriction digests on an agarose gel. The size-selected DNA was ligated and packaged *in vitro* into the λ phage

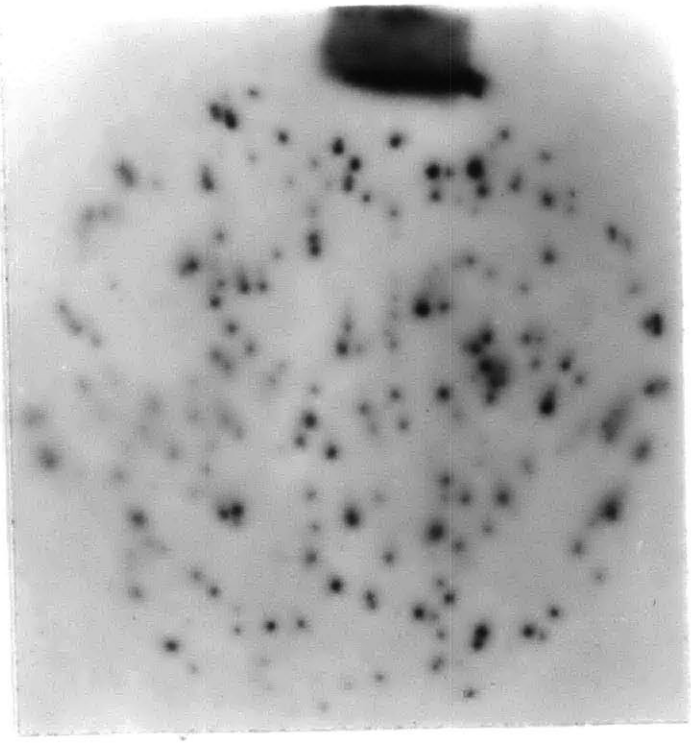
vector CH4A*. The libraries constructed from lymphocyte cell lines GM4207 and GM4219 were designated HDA and HDB, respectively. The number of independently derived phage clones for each library was $3.2X10^5$ for HDA and $3.4X10^4$ for HDB. The CH4A* vector contains the lac 5 substitution which carries the E. coli gene for β-galactosidase (lac z). When the chromogenic substrate XG is included in the plating medium, phage carrying lac 5 give blue plaques. This provides a useful indicator of successful substitution of human DNA for the dispensable fragments of the CH4A* vector. The HDA library and the HDB library gave 95% and 80% colorless (recombinant) phage plaques, respectively. Libraries were hybridized to ^{32}P-labelled HeLa DNA (S.A. $1X10^8$ cpm/μg DNA) in order to estimate the percentage of recombinant phage clones containing human repetitive DNA sequences. It would be expected that a high proportion of recombinant clones would hybridize to such a probe reflecting the frequency and interspersion of repeated DNA sequences in human DNA fragments of the 13 kb to 17 kb size range. Greater than 70% of the recombinant phage clones from libraries HDA and HDB hybridized to the HeLa probe (Figure 5, panel I).

Isolation of Matching Clone Sets from Libraries HDA & HDB

The following strategy was employed to isolate homologous recombinant phage clones (matching clones) from both libraries. Random clones were picked from both libraries and phage DNA prepared. DNA was digested with a variety of restriction endonucleases in combination with Eco RI in order to obtain discrete human DNA fragments separated from λ phage DNA. DNA was fractionated on agarose gels and analyzed by blot



HDA



HDB

Figure 5: Hybridization of Total Human DNA Probe to Human Recombinant DNA Libraries HDA & HDB.

Approximately 1,000 individual phages of the HDA library were plated on a bacterial lawn; approximately 500 phages from the HDB library were plated. Filters were made from both plates and hybridized to nick-translated total human DNA probe.

hybridization (21) to HeLa probe ($1X10^8$ cpm/μg DNA). Single copy human DNA fragments were identified and selected for use as hybridization probes to screen the libraries for matching clones. A random clone (HD1B) selected from library HDB contains a 3.0 kb human single-copy DNA fragment when digested with Eco RI plus Sac I. When this fragment was purified and used as a hybridization probe to screen 10,000 phage clones of the HDA library, 10 phage clones showed intense hybridization. Three of these clones were plaque purified and DNA prepared for comparative restriction endonuclease analysis. Similarly, a random recombinant phage clone (HD2A) from the HDA library contains an 8 kb human single-copy DNA fragment when digested with Eco RI and Pvu II. When this fragment was used as a hybridization probe to screen 10,000 clones from both the HDA and HDB libraries, 15 phage clones from the HDA library and 5 phage clones from the HDB library showed intense hybridization. These results are consistent with the relative complexities of the libraries. One of the hybridizing clones (HD2B) from the HDB library, and an additional clone from the HDA library were plaque purified and prepared for comparative restriction endonuclease analysis.

Comparative Restriction Endonuclease Cleavage Site Analysis

Two sets of matching clones HD1A, HD1B, and HD2A, HD2B were analyzed with 22 restriction endonucleases. A human DNA insert size of 13.8 kb for HD1A & HD1B, and 14.8 kb for HD2A & HD2B, was determined by Eco RI digestion of the phage clones using Hind III-digested λcI857 as molecular weight standards. Restriction fragment length polymorphisms (RFLPs) were identified by comparing restriction digests of matching

clone DNAs electrophoresed in adjacent lanes of ethidium bromide stained gels. In a typical experiment (Figure 6), a number of phage clones from a set of matching clones were digested with the same restriction endonucleases. The restriction fragment patterns were compared and any restriction fragment length variations noted. The total number of human DNA fragments generated by a given endonuclease (determined by subtracting from the total number of restriction fragments those that were phage DNA fragments) was used to determine the number of restriction sites examined. As illustrated in Figure 6, three phage clones, two from the HDA library and one from the HDB library show identical restriction fragment patterns for four different restriction endonucleases. The number of restriction sites sampled by these four enzymes is indicated in Table II.

The results of the restriction endonuclease analysis for two sets of matching clones are summarized in Table II. In addition to the enzymes listed, patterns generated by Pst I, Bgl I, and Hae II were also analyzed for RFLPs. Although no RFLPs were observed for these enzymes, the number of cleavage sites in human DNA could not be accurately determined since no restriction map of these enzyme sites in CH4A* exists. Clones HD2A and HD2B shared identical cleavage sites for all 22 endonucleases (Table II). Clones HD1A and HD1B exhibited four differences in restriction fragment length when digested with Alu I, Hae III, and Taq I. An additional fragment is present in clone HD1B when digested with Alu I that is not present in HD1A. Hae III produces similar numbers of fragments in both clones; however, as shown in Figure 7, two restriction fragment length differences are observed that result in pairs of restriction fragments that differ by 40 bp and 60 bp.

Table II. Comparison of Restriction Endonuclease Recognition Sites in Matching Clones HD1A & HD1B, HD2A & HD2B

Enzyme	No. of Enzyme Recognition Sites in Clones		Total No. of Base Pairs in Recognition Sites	
	HD1A & HD1B	HD2A & HD2B	HD1A & HD1B	HD2A & HD2B
Eco R1	2	2	12	12
Bgl II	2	4	12	24
Hind III	4	3	24	18
Bam HI	0	1	0	6
Kpn I	1	1	6	6
Sac I	2	3	12	18
Pvu II	2	1	12	6
Xho I	0	0	0	0
Xba I	9	3	54	18
Sac II	0	5	0	30
Msp I	0	>1	0	>4
Hae III	>13	>11	>52	>44
Rsa I	>18	>15	>72	>60
BstNI	5	6	25	30
Hinf I	>14	>11	>70	>55
Sal I	0	0	0	0
Hha I	0	1	0	4
Alu I (HD1A)	>15	>10	(HD1A) >60	>40
(HD1B)	>16		(HD1B) >64	
Taq I (HD1A)	10	8	(HD1A) 40	32
(HD1B)	11		(HD1B) 44	
Total No. of Base Pairs in Enzyme Recognition Sites:			>459	>407

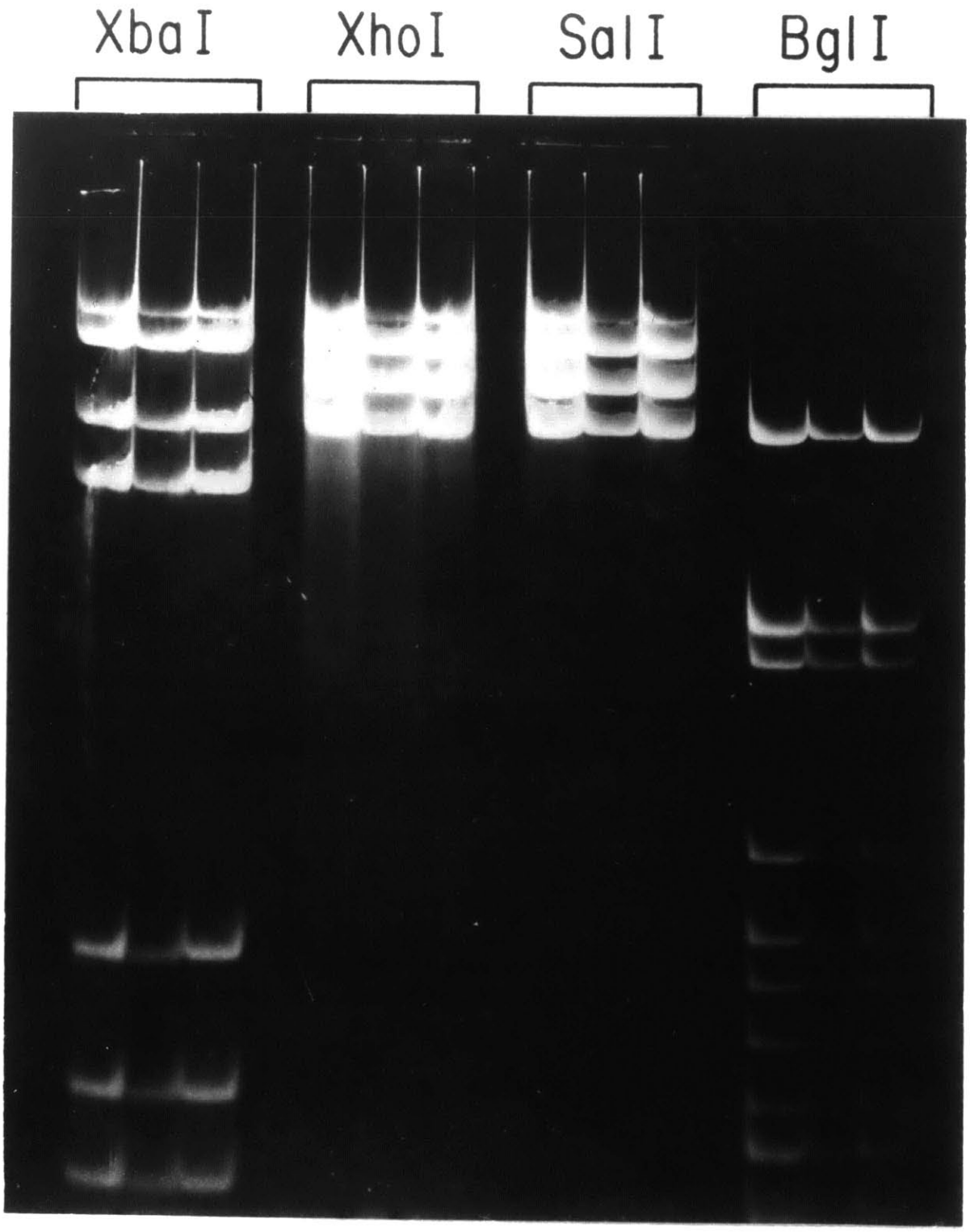

XbaI
XhoI
SalI
BglI

Figure 6: Comparison of Restriction Fragment Patterns of Matching Clones HD2A & HD2B.

DNA from three phage clones, two from the HDA library and one from the HDB library was digested with Xba I, Xho I, Sal I, and Bgl I. Restriction digests were fractionated on a 1% agarose gel. DNA was visualized by staining with ethidium bromide.

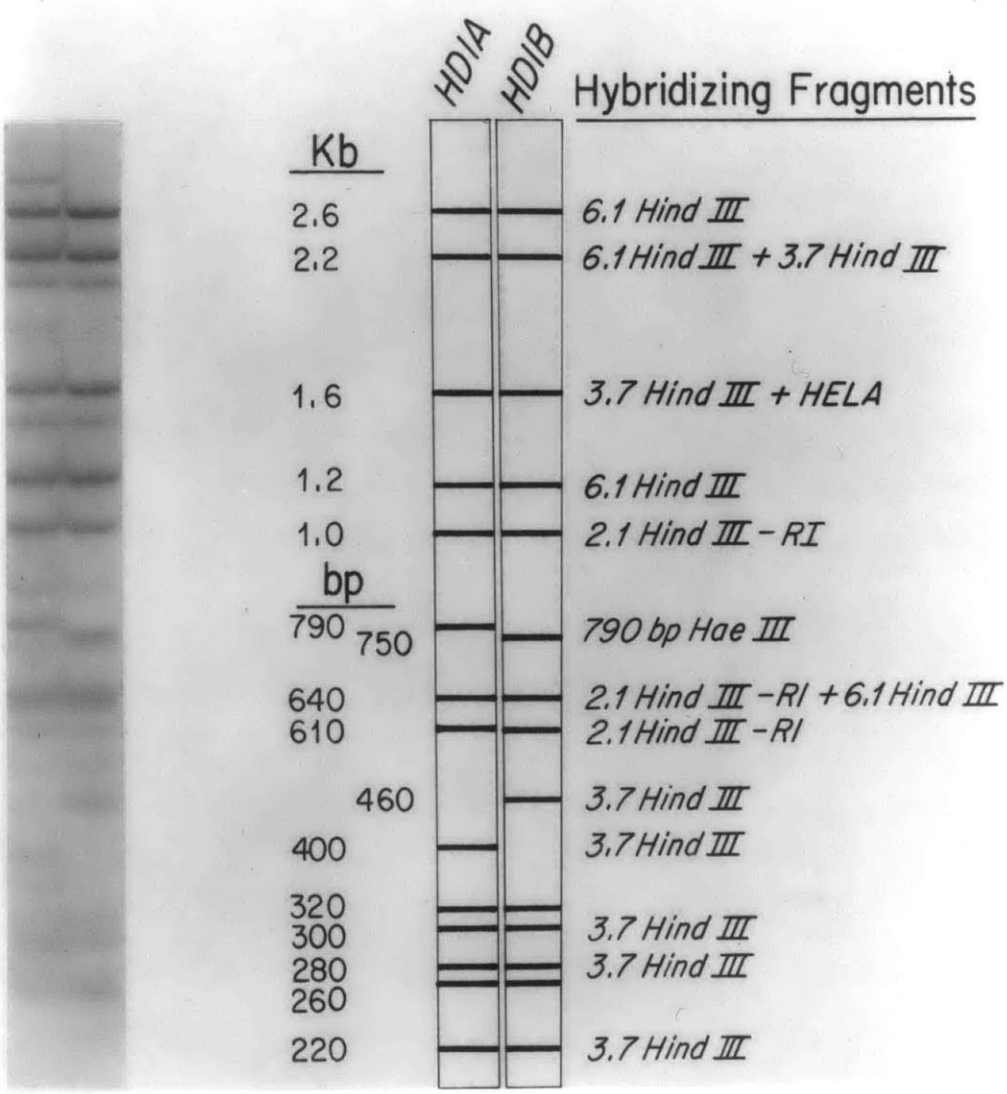

HDIA
HDIB
Hybridizing Fragments
Kb
2.6
2.2
1.6
1.2
1.0
bp
790
750
640
610
460
400
320
300
280
260
220
6.1 Hind III
6.1 Hind III + 3.7 Hind III
3.7 Hind III + HELA
6.1 Hind III
2.1 Hind III - RI
790 bp Hae III
2.1 Hind III - RI + 6.1 Hind III
2.1 Hind III - RI
3.7 Hind III
3.7 Hind III
3.7 Hind III
3.7 Hind III
3.7 Hind III

Figure 7: Hae III Digestion of Matching Clones HD1A and HD1B and Schematic of Hybridization to HeLa and HD1A Subfragment Probes.

DNA from two phage clones, one (HD1A) from the HDA library and one (HD1B) from the HDB library, was digested with Hae III, fractionated on a 2% agarose gel, and visualized by staining with ethidium bromide. The gel was transferred to nitrocellulose and hybridized to nick-translated total human DNA (HeLa cell DNA), and nick-translated subfragments of clone HD1A (see restriction map) as indicated by schematic.

(Those bands that are visible in the gel depicted in Figure 7 but not represented in the schematic are due to contamination of the purified human DNA inserts with phage DNA. This was confirmed by hybridization to lambda DNA, data not shown.) Taq I digestion produces a 3 kb fragment in HD1A that is not present in HD1B. Instead, 2 kb and 1 kb fragments are generated in HD1B that are not present in HD1A (Figure 8).

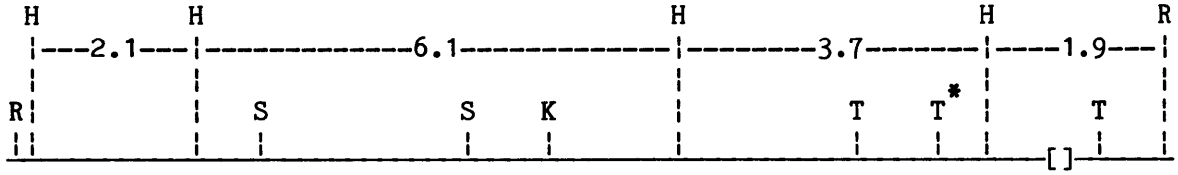
A total of 866 bp of DNA sequence was screened for restriction endonuclease cleavage site variants in the two sets of matching clones (Table II). This is a minimum estimate of the number of base pairs screened since fragments less than 300 bp on 1% gels or less than 150 bp on 2% gels would not be scored under the conditions used. Assuming that the presence of or absence of a restriction endonuclease recognition site is the result of a single base pair change, a total of at least 4 out of 866 bp screened show polymorphic variation. All 4 restriction endonuclease cleavage site variants were found in one set of matching clones (HD1A & HD1B). The other set of matching clones (HD2A & HD2B) shared identical DNA sequence in all restriction sites examined, representing 407 bp of DNA sequence. The frequency of polymorphic variation in the 866 bp of DNA sequence examined using this method is 0.0046.

Restriction fragment length differences of as little as 40 bp can be easily resolved on 2% gels. The smallest fragment that can be scored for RFLP is 150 bp on 2% gels and 300 bp on 1% gels.

Localization of Hae III RFLPs

In an attempt to localize the Hae III RFLPs, single-copy DNA

subfragments of HD1A were identified and hybridized to Hae III-digested HD1A and HD1B. The location of the cleavage sites within the 13.8 kb insert of HD1A for the enzymes Bam HI, Sac I, Hind III, and Kpn I were established. The restriction enzyme map is shown in Figure 8. The location of repetitive DNA sequences in this set of clones had been previously assigned to a single 1.9 kb Hind III-Eco RI fragment by blot hybridization to ^{32}P-labelled HeLa probe during the search for single-copy probes for matching clones. HeLa DNA and the three single-copy subfragments of HD1A (2.1 kb, 6.1 kb, and 3.7 kb) generated by digestion with Hind III and Eco RI were hybridized sequentially to Hae III-digested HD1A and HD1B. The results of the hybridizations are illustrated by the schematic of Figure 7. A single 1.6 kb Hae III fragment common to both HD1A and HD1B hybridized to the HeLa probe. Neither of the two Hae III RFLPs were detected by hybridization to the 6.1 kb and 2.1 kb subfragments. One of the Hae III RFLPs was localized to the 3.7 kb subfragment by hybridization of that subfragment to a 400 bp fragment of HD1A and a 460 bp fragment of HD1B. The lack of detectable hybridization of the other Hae III RFLP (790 bp fragment of HD1A and 750 bp fragment of HD1B) to any of the single-copy DNA subfragments of HD1A suggested that this RFLP should be located somewhere within the 1.9 kb subfragment that contains the repetitive DNA sequence. Hae III digestion of the 1.9 kb subfragment of HD1A produced a 790 bp single copy DNA fragment. When this fragment was used as a hybridization probe, the 790 bp fragment of HD1A and the 750 bp fragment of HD1B were observed to hybridize thus confirming the localization of this Hae III RFLP to the 1.9 kb subfragment. These results are consistent with the existence of two independent Hae III RFLPs in the



1 kb

Figure 8: Restriction Endonuclease Cleavage Site Map of Clone HD1A. The restriction map sites for Eco RI, Bam I, Hind III, Sac I, and Kpn I were established by analyzing restriction digests of Clone HD1A produced by combinations of the above enzymes. The location of repetitive DNA within the clone was deduced from hybridization to nick-translated HeLa DNA probe (data not shown). [ ] indicates repetitive DNA sequences. R, Eco RI; S, Sac I; K, Kpn I; H, Hind III; T, Taq I. The Taq I sites are not precisely mapped but are deduced from the hybridization results illustrated in Figure 9. *T indicates a presence/absence for a Taq I cleavage site.

matching clones HD1A and HD1B. The results do not distinguish between the generation of the Hae III RFLPs by single base changes which add or remove a base pair from the 4 bp Hae III recognition site (GGCC) or by small deletions or insertions of 40 bp or 60 bp of DNA sequence since fragments of this size would not be scored under the conditions used.

The above experiments served to localize the Hae III RFLPs to specific areas within the cloned segments of DNA. Additional experiments were designed to confirm the pattern of RFLP in genomic DNA of the individuals from whom the libraries were made and to test the inheritance of the Hae III RFLPs in the Huntington's Disease pedigree. The 3.7 kb and the 790 bp single copy DNA subfragments of HD1A that were used as hybridization probes for the Hae III RFLPs were hybridized to Hae III-digested human DNA. A major difficulty in the hybridization of single-copy DNA fragments in this relatively small size range resulted in only a very weak hybridization of the 3.7 kb probe to the 460 bp and 400 bp Hae III fragments of GM4219 and GM4207, respectively. These results confirm the existence of at least one of the Hae III RFLPs in the human genome for two individuals and are consistent with the pattern observed with cloned DNA fragments isolated from libraries of DNA made from those individuals.

Characterization of Taq I RFLP

An approach similar to the one used to localize the Hae III RFLPs was used to characterize the Taq I RFLP observed in clones HD1A and HD1B. A single-copy DNA subfragment of HD1A was hybridized to Taq I-digested HD1A and HD1B in order to specify the area of the cloned DNA

containing the Taq I RFLP. The restriction fragment pattern of Taq I-digested HD1A and HD1B DNA in the ethidium bromide stained gel and the results of hybridization of the HD1A 3.7 kb subfragment and HeLa DNA are shown in Figure 9. The HeLa probe hybridized to the 3 kb fragment of HD1A and the 2 kb fragment of HD1B. This result localizes the repetitive DNA sequence to the 3 kb and 2 kb Taq I-digested DNA fragments of HD1A and HD1B, respectively. This result further suggests that the Taq I RFLP should be localized to the 3.7 kb Hind III subfragment of HD1A. This localization is confirmed by the hybridization of the 3.7 kb probe to the 3 kb fragment of HD1A and both the 2 kb and 1 kb fragments of HD1B as shown in Figure 9. These results demonstrate the presence of a Taq I recognition site in the 3.7 kb subfragment of HD1B that is not present in HD1A. Taq I recognizes a 4 base pair sequence, TCGA. The most likely explanation for the Taq I RFLP is a single base change due to a substitution or modification that creates a Taq I recognition site in HD1B or eliminates a recognition site in HD1A.

Inheritance of the Taq I RFLP

In order to obtain a suitable amount of the 3.7 kb Hind III single-copy subfragment of HD1A to be used as a hybridization probe for the Taq I RFLP, the 3.7 kb fragment was subcloned into the λ phage vector Charon 21A. Results of an *in situ* plaque hybridization for Charon 21A phage that incorporated the 3.7 kb fragment are shown in Figure 10. An intensely hybridizing plaque containing the 3.7 kb Hind III single-copy subfragment of HD1A was selected and designated HD1A-H4.

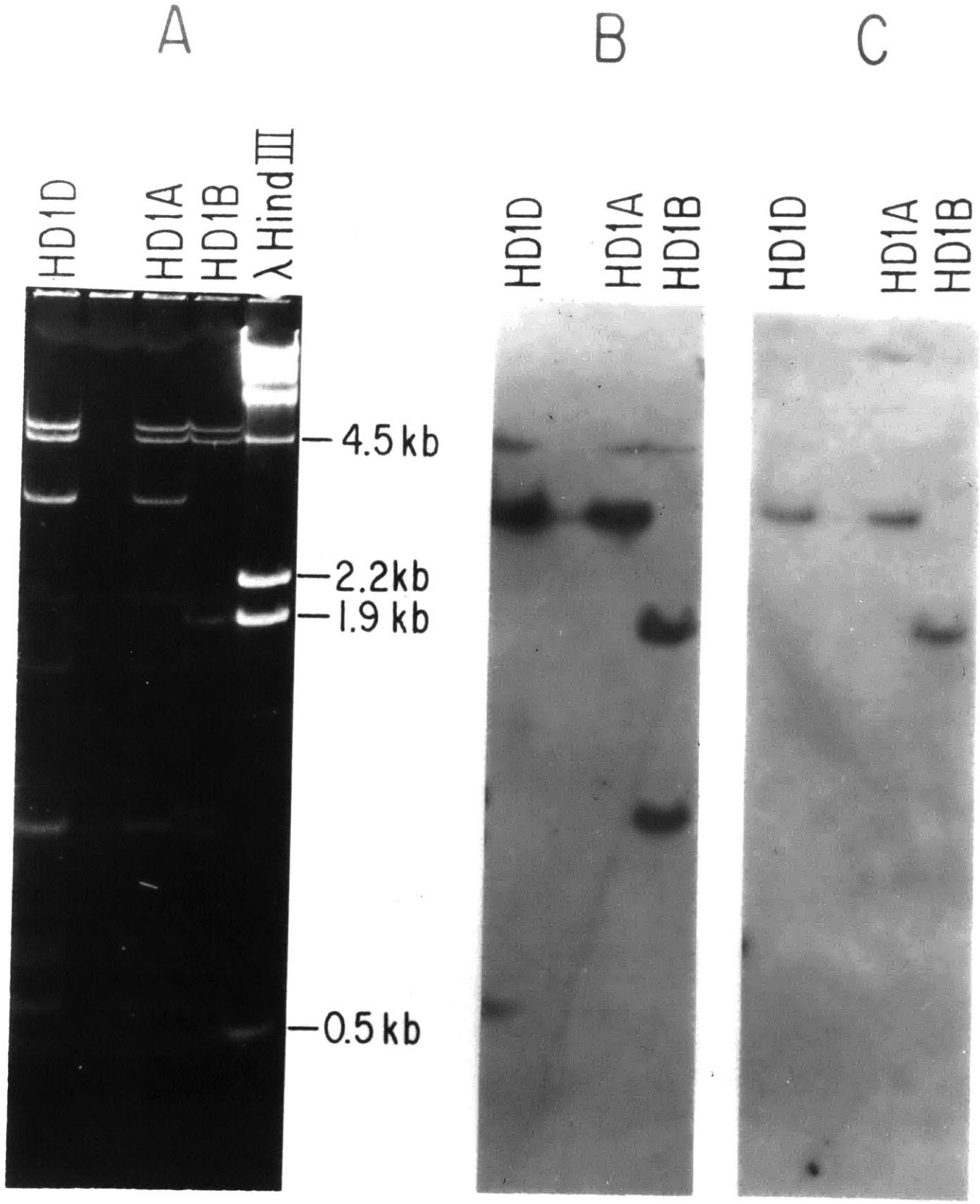

A
B
C
HD1D
HD1A
HD1B
λ HindIII
4.5 kb
2.2kb
1.9 kb
0.5 kb
HD1D
HD1A
HD1B
HD1D
HD1A
HD1B

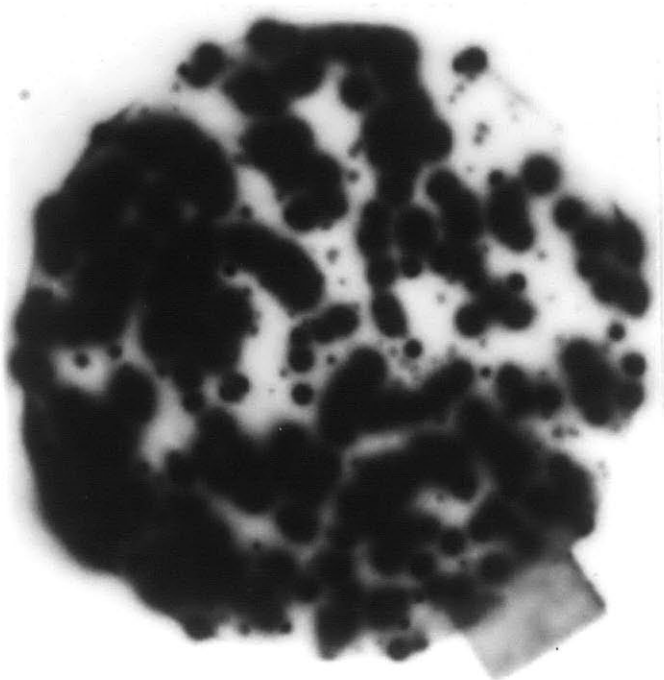
Figure 9: Taq I Digestion of Clones HD1A, HD1B, & HD1D and Hybridization to HeLa and HD1A-H4 Probes.

Panel A- Two phage clones (HD1A & HD1D) from the HDA library and one clone (HD1B) from the HDB library, were digested with Taq I and fractionated on a 2% agarose gel. DNA was visualized by staining with ethidium bromide. Hind III digested λcI857Sam7 DNA was run in an adjacent lane as approximate size markers.

Panel B- The gel described in panel A was transferred to nitrocellulose and hybridized to nick-translated HD1A-H4 probe.

Panel C- A similar experiment to panel B was performed except that the hybridization probe was nick-translated total human DNA (HeLa cell DNA).

II



I

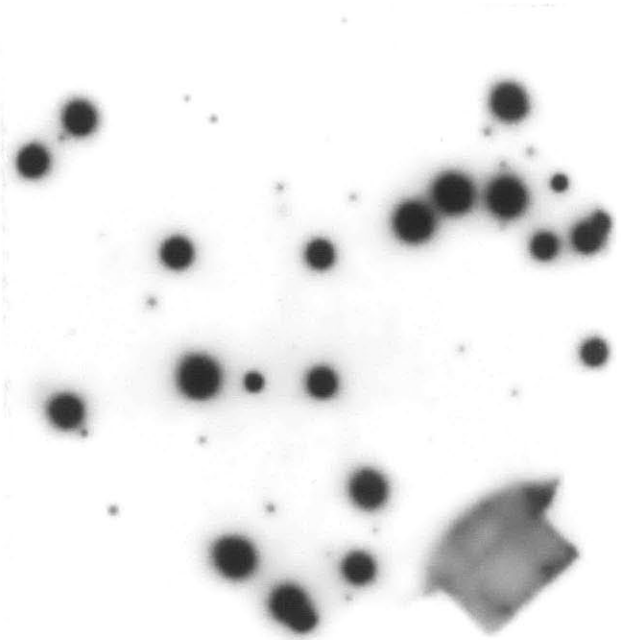


Figure 10: Subcloning of HD1A Subfragment (HD1A-H4) into Charon 21A. Purified 3.7 kb single-copy DNA subfragments of clone HD1A were ligated to end fragments of Hind III digested Charon 21A λ phage vector DNA. The ligation mixture was packaged in vitro. Recombinant phage clones were screened by plating approximately 100 phage clones (panel I), or 500 phage clones (panel II) on bacterial lawns. Filters were made and hybridized to nick-translated HD1A 3.7 kb single-copy subfragment DNA probe.

The 3.7 kb human insert of HD1A-H4 was purified and used as a hybridization probe to analyze the restriction fragment length pattern of the Taq I-digested DNA of 22 individuals. The results of the hybridization for 10 individuals is shown in Figure 11. The restriction fragment length patterns exhibited by individuals 4207 and 4219 are consistent with the patterns observed in the matching clones HD1A and HD1B which contained the cloned homologous DNA fragments of those individuals. Two alleles are defined by hybridization of the 3.7 kb probe to a restriction fragment of either 3 kb or restriction fragments of 2 kb and 1 kb. Individual 4207 is homozygous for the 3 kb fragment while individual 4219 is heterozygous for the 3 kb, 2 kb, and 1 kb fragment lengths. The pattern of inheritance of restriction fragment lengths for 22 individuals is shown in Table III. A pedigree showing the familial relationships of those individuals from within a Huntington's Disease pedigree is shown in Figures 3 & 4. 11 individuals are homozygous for the 3 kb fragment length while the other 11 individuals are heterozygous for the 3 kb, 2 kb, and 1 kb fragment lengths (Figure 4). The restriction fragment lengths observed are consistent with their inheritance as Mendelian alleles through 3 generations. No correlation is observed between the inheritance of either Taq I fragment length pattern and the autosomal dominant genetic disease locus responsible for Huntington's Disease.

9 unrelated individuals from the above panel of 22 individuals were used to determine the allele frequencies of the Taq I RFLP. 6 individuals were heterozygous for the 3 kb, and 2 kb, 1 kb fragment lengths, while 3 individuals were homozygous for the 3 kb fragment length. This distribution gives a 3 kb fragment length allele frequency of 0.7.

Table III. Hybridization of EcoR1 and Taq 1 Fragments to Clones G6 and HD1A-H4.

	Clones	
Individuals	G6	HD1A-H4
GM4201	Aa	Bb
GM4203	aa	
GM4207	Aa	BB
GM4209	AA	BB
GM4211	AA	
GM4213	Aa	
GM4217	Aa	BB
GM4219	AA	Bb
GM4221	Aa	Bb
GM4223	Aa	Bb
GM4225		BB
GM4227	AA	Bb
GM4229	Aa	BB
GM4233	AA	BB
GM4189	Aa	Bb
GM4191	Aa	Bb
GM4193	AA	
GM4195	Aa	BB
GM4197	Aa	BB
GM4199	Aa	
Gus 5	AA	Bb
Gus 13		Bb
Gus 42	AA	Bb
Gus 50	Aa	BB
86B	Aa	Bb
114A	Aa	BB
GM2064		BB

A= Upper EcoR1 fragment
B= Upper Taq 1 fragment
a= Lower EcoR1 fragment
b= Lower Taq 1 fragments

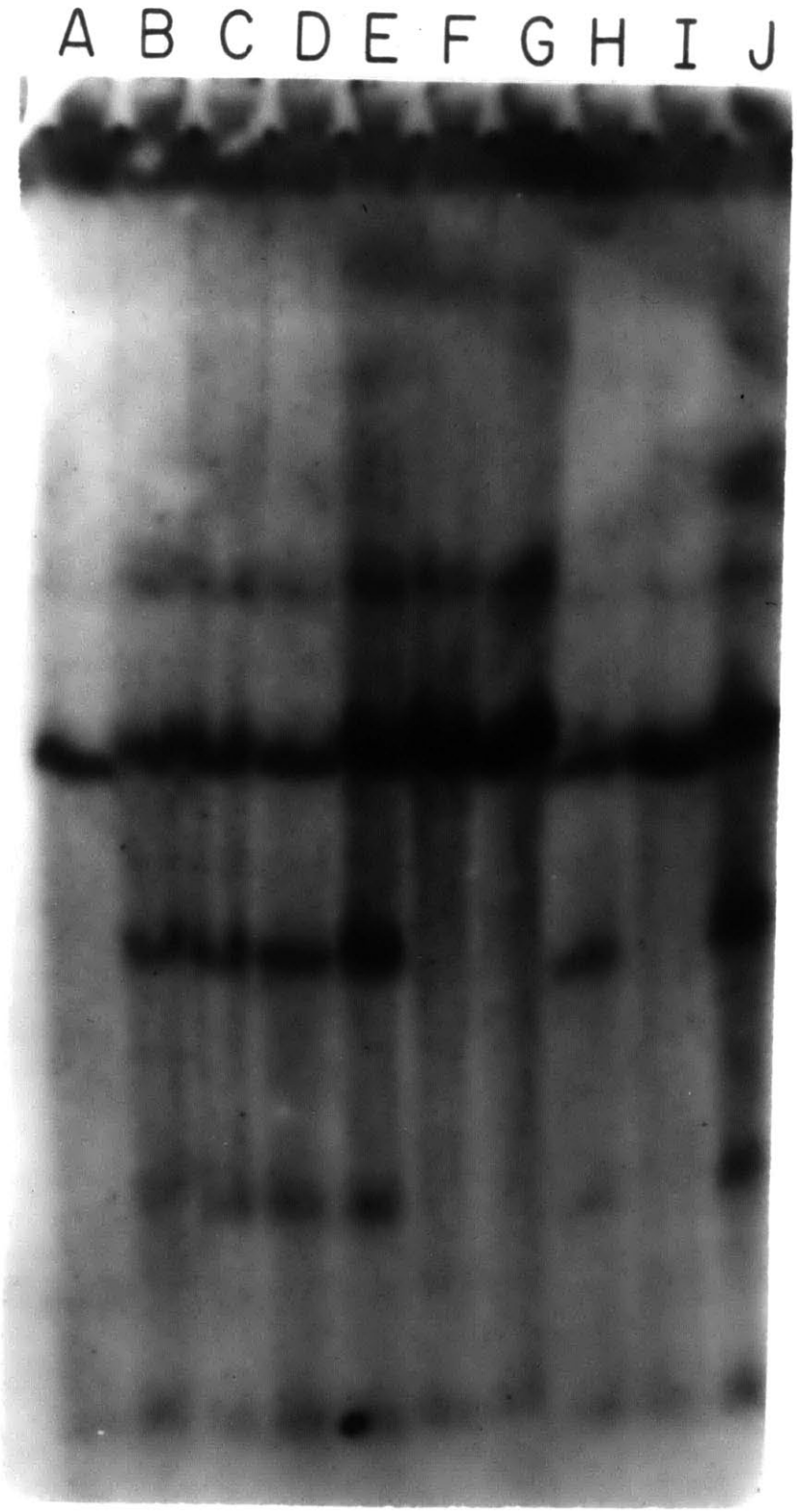

A B C D E F G H I J

Figure 11: Hybridization of HD1A-H4 Probe to Taq I Digested Cell DNAs. DNA extracted from the indicated cell line was digested with Taq I, fractionated on a 2% agarose gel, transferred to nitrocellulose and hybridized to ^{32}P-labelled HD1A-H4 DNA.

Lanes: A, GM4207; B, GM4219; C, GM4189; D, GM4191; E, GM4227; F, GM4233; G, GM4197; H, 4201; I, GM4195; J, GM4233.

CHAPTER 3

MIDDLE REPETITIVE DNA PROBES FOR DNA POLYMORPHISMS

RATIONALE

There are several advantages to using human middle repetitive DNA (1,000-2,000 copies/haploid genome) as probes for RFLPs. The major advantage is the ability to screen, with one hybridization probe, a large number of restriction endonuclease cleavage sites in a single experiment. This advantage can also be a major limitation since the restriction fragment pattern becomes incredibly complex with considerable loss of resolution of individual fragments when analyzed by the conventional Southern hybridization method (21). This study employs the following strategy to achieve resolution of individual hybridizing fragments: The DNA from two individuals to be compared for RFLPs is first digested with a restriction endonuclease and fractionated by agarose gel electrophoresis. Specific size fractions of DNA are then cut out of the gel and electroeluted from the agarose. Each DNA size fraction is then digested with a second restriction endonuclease. The resulting restriction fragment digests are fractionated in agarose. For ease of comparison, equivalent size fractions of restriction digests from each individual are electrophoresed in adjacent lanes of the gel.

Two middle repetitive DNA hybridization probes were used to analyze the restriction digests of two individuals. Three different restriction

enzyme combinations were used. A total of 7 variant restriction fragment length patterns were observed for an estimated total of 6400 bp of DNA sequence.

MATERIALS AND METHODS

Digestion, Electrophoresis, Transfer, and Hybridization of DNA

300 µg aliquots of DNA isolated from lymphocyte cell lines GM4207 and GM4219 (Human Genetic Mutant Cell Repository, Camden, New Jersey) by methods previously described were digested with 450 units of Bam HI or 450 units of Eco RI at a DNA concentration of 1.2 mg DNA/ml. Restriction digests of GM4207 and GM4219 were loaded onto separate halves of a 1% agarose gel and electrophoresed at 40 volts for 24 hrs. DNA was visualized by staining with ethidium bromide. The gel was sliced horizontally in 6 mm slices beginning with the first visible high molecular weight fraction of DNA. DNA from the top 8 fractions of both GM4207 and GM 4219 was electroeluted from the agarose by methods described in Chapter 2. After phenol-chloroform extraction and precipitation with ethanol, the DNA was dried and resuspended in 50 µl of TE. 20 µl aliquots of each Eco RI-digested DNA sample was then digested with either Sac I (80 units) or Taq I (60 units). 20 µl aliquots of Bam HI-digested DNA fractions were digested with Eco RI (75 units). DNA samples were applied to 0.8% agarose gels (Eco RI-Sac I-digested DNA or Bam HI-Eco RI-digested DNA) or to 2.0% agarose gels (Eco RI-Taq I-digested DNA). DNA samples from parallel gel slices that were

digested with the same combination of enzymes were run in adjacent lanes of an agarose gel and electrophoresed at 40 volts for 18 hrs. The gels were transferred to nitrocellulose by the procedure of Southern (21) as described in detail in Chapter 1. Hybridization was carried out as specified in Chapter 1.

Hybridization Probes

Two middle repetitive DNA fragments were used as hybridization probes. A 2.2 kb Eco RI-Pst I subclone of the A36 plasmid (designated A36FC) characterized by Duncan et al. (30) was used as a hybridization probe. This DNA fragment is located in the β-globin region of human chromosome 11 , 3.5 kb upstream from the $^{G}\gamma$-globin gene. A middle repetitive DNA clone (designated NAR-1 for non-Alu repetitive) isolated from the human fetal liver DNA library constructed by Maniatis (22) was provided by J. Gusella. Both repetitive DNA elements can be characterized as interspersed middle repetitive DNA and are not members of the Alu family, the predominant class of interspersed, repetitive DNA in the human genome. A36FC plasmid or NAR-1 phage DNA was labelled with ^{32}P to a specific activity of $1.0X10^{8}$ cpm per/μg DNA by the procedure of Rigby et al. (25).

RESULTS

Two middle repetitive human DNA fragments were used as hybridization probes to screen restriction endonuclease digested DNA of

two individuals for restriction fragment length differences. Specific size classes of restriction digests were compared by Southern blot analysis after digestion with a combination of two restriction endonucleases. The first restriction digest allowed isolation of specific size classes of DNA fragments by excision and electroelution of DNA following electrophoresis in agarose. The second restriction digest of similar size classes of electroeluted DNA fragments from two individuals further reduced the complexity of DNA fragments contained in the digest, thus allowing resolution of DNA fragments that hybridized to middle repetitive DNA probes when analyzed by the Southern transfer method (21). Restriction fragment length differences could be identified by the presence/absence of DNA fragments or by differences in mobility of DNA fragments when restriction digests of specific size classes of DNA were compared by applying to adjacent lanes of an agarose gel the same size class of restriction digests from two individuals. Restriction fragment length differences identified in this manner could result from the elimination or addition of a restriction endonuclease cleavage site flanking the middle repetitive hybridizing sequence in the genome. Alternatively, DNA rearrangements such as insertions or deletions flanking the hybridizing sequences would also result in restriction fragment length differences. Similarly, modifications of the DNA sequence within the middle repetitive DNA itself could result in restriction fragment length differences due to insertions or deletions of blocks of DNA or in the generation of a pair of new restriction fragments if a base pair substitution resulted in the creation of a new restriction endonuclease cleavage site.

DNA isolated from lymphocyte cell lines GM4207 and GM4219 was

digested with Eco RI. Restriction digests were fractionated by agarose gel electrophoresis and specific size classes of DNA fragments were excised and electroeluted from the agarose gel. Aliquots from 8 similar size classes of Eco RI-digested DNA from each individual were digested with Sac I and applied to adjacent lanes of an agarose gel, electrophoresed, and transferred to nitrocellulose according to the procedure of Southern (21). The filter was hybridized to $15X10^6$ cpm of ^{32}P-labelled A36FC DNA (S.A. $2.0X10^8$ cpm/μg DNA). The results of the hybridization of A36FC to 6 pairs of restriction digests are illustrated in Figure 12. It is apparent that a large number of hybridizing fragments can be resolved by this method. A total of approximately 120 DNA fragments can be resolved for each individual (Figure 12). 164 hybridizing DNA fragments can be resolved for each individual for all 8 pairs of restriction digests (two pairs not shown in Figure 12). A total of 4 differences in the presence/absence of DNA fragments between pairs of adjacent restriction digests were scored. Two of these differences are illustrated in Figure 12 by the presence of two DNA fragments in 4A not present in 4B, in the upper third of the autoradiograph. The number of unique hybridizing DNA fragments is slightly overestimated by this method since some DNA fragments are shared by two pairs of restriction digests. These DNA fragments can be accounted for by their presence near the ends of the agarose gel slices used to fractionate the first restriction digests into size classes. The intensely hybridizing DNA fragment in the lower third of Figure 12 that is common to all restriction digests most likely represents an internal Eco RI or Sac I endonuclease cleavage site in the genomic middle repetitive DNA hybridizing to the A36FC probe.

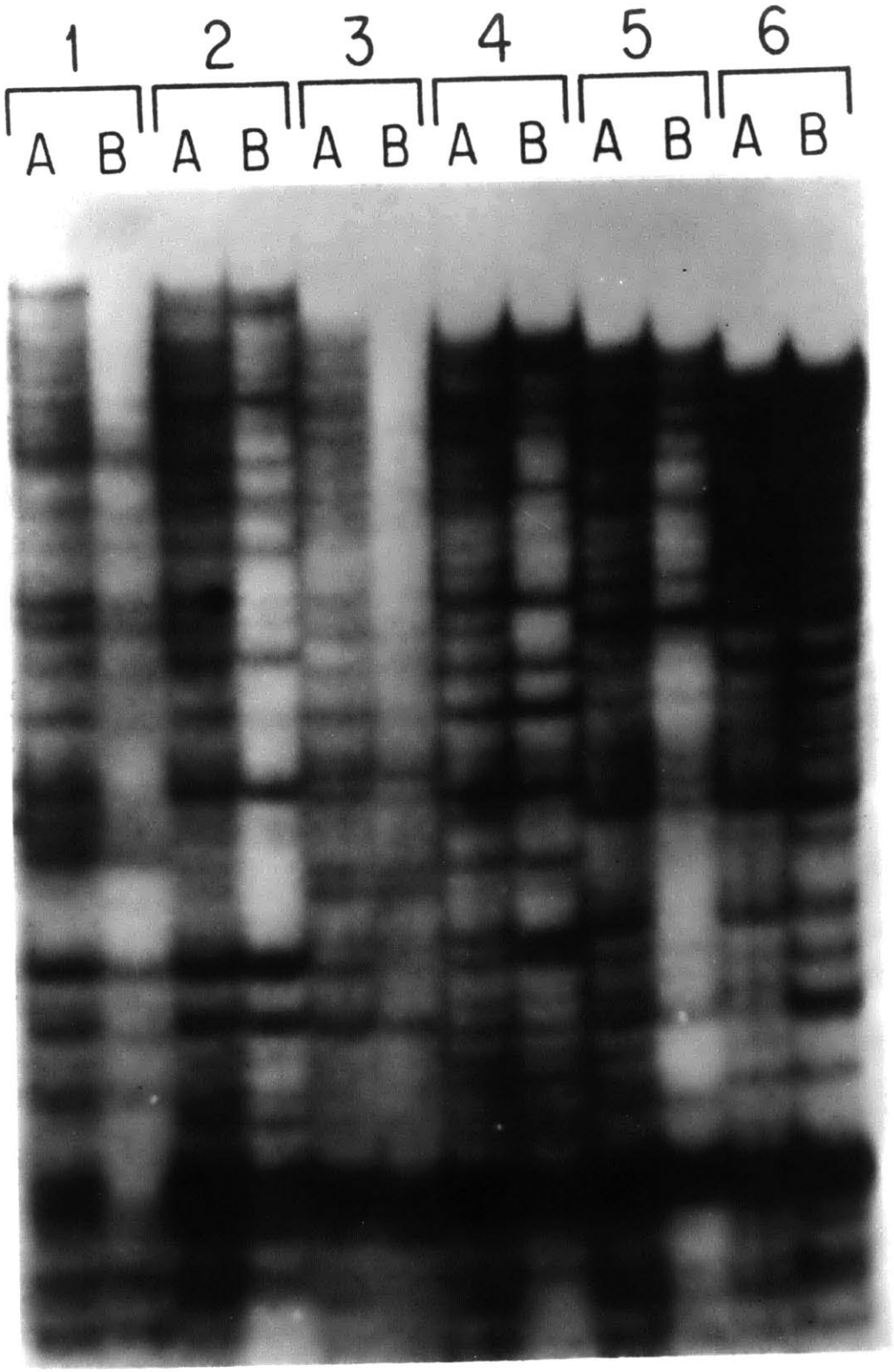

1
2
3
4
5
6
A B A B A B A B A B A B

Figure 12: Hybridization of A36FC Middle Repetitive DNA Probe to Eco RI-Sac I Digested Cell DNAs.

DNA extracted from the GM4219 and GM4207 cell line was digested with Eco RI and fractionated on separate halves of a 1% agarose gel. The gel was sliced horizontally and DNA was electroeluted from each 6mm gel slice. DNA from parallel gel slices of GM4207 and GM4219 (equivalent size fractions of DNA) were digested with Sac I, fractionated on a 1% agarose gel, transferred to nitrocellulose, and hybridized to nick-translated A36FC probe.

Lanes: A, GM4207; B, GM4219 1-6 designates pairs of equivalent size fractions of restriction digests.

The total copy number of DNA sequences homologous to the A36FC probe in the human genome can be roughly estimated by this method. Assuming that the amount of DNA analyzed by this method represents approximately 10% of the genome, and allowing for the redundancy in scoring DNA fragments that result from either internal cleavage sites in the genomic repetitive DNA sequences or from the sharing of DNA fragments by two pairs of restriction digests, the copy number of DNA sequences homologous to A36FC probe in the human genome is approximately 1400.

Restriction fragment length differences identified in this manner can result from the creation or elimination of a restriction endonuclease cleavage site for either of two endonucleases (Eco RI or Sac I), or from DNA rearrangements such as insertions or deletions. The number of base pairs of DNA sequence analyzed by this method can be estimated, assuming each hybridizing DNA fragment represents the sampling of two restriction endonuclease cleavage sites. This assumption is valid if the distribution of the hybridizing repetitive DNA sequences in the human genome is such that repetitive DNA sequences are flanked by long tracts of single-copy DNA. The relatively low copy number estimated above and the distribution and frequency of A36FC in the β-globin region of the human genome (30) are not inconsistent with this assumption. Since both Eco RI and Sac I recognize a 6 bp DNA sequence, 164 hybridizing DNA fragments detected for each of GM4219 and GM4207 would represent 328 cleavage sites or 1,968 bp of DNA sequence per individual. If the 4 restriction fragment length differences scored using this method are the result of single base pair changes then a total of 4 out of 1,968 bp show potential polymorphic variation.

Additional experiments were designed to examine a number of different restriction endonuclease combinations for both the first and second restriction digests and to include another middle repetitive DNA hybridization probe. A λ phage clone containing middle repetitive DNA was isolated from a library of human DNA fragments and designated NAR-1. The results of a series of hybridizations of A36FC and NAR-1 to restriction digests of two individuals are summarized in Table IV. No restriction fragment length variation between individuals was observed when DNA was digested with a combination of Bam HI-Eco RI and hybridized to A36FC. A total of 161 hybridizing DNA fragments representing approximately 322 cleavage sites or 1932 bp DNA sequence per individual were scored. Three restriction fragment length differences from a total of 128 hybridizing DNA fragments (GM4207) or 129 DNA fragments (GM4219) were found when A36FC was hybridized to Eco RI-Taq I restriction digests. Assuming that Taq I accounts for the majority of restriction cleavage sites screened by this method, 129 hybridizing DNA fragments represent 258 cleavage sites or 1290 bp of DNA sequence, since Taq I recognizes a 4 bp DNA sequence. A variation of 3 bp out of 1290 is observed, assuming the restriction fragment length differences are the result of single base pair changes. Hybridization of Eco RI-Sac I restriction digests to NAR-1 probe resulted in the same number of hybridizing DNA fragments (101) for both individuals. No restriction fragment length differences scored. This represents the screening of 202 cleavage sites, or 1212 bp of DNA sequence, with no variation observed.

Out of a total of approximately 1100 restriction endonuclease cleavage sites examined, representing 6400 bp of DNA sequence, all

Table IV. No. of Hybridizing DNA Fragments in Individuals GM4207 & GM4219 Detected by Hybridization to Repetitive DNA

Restriction Endonuclease Combinations	Hybridization Probe: A36FC		No. of Differences
	GM4207	GM4219	
Eco RI-Sac I	164	164	4
Bam HI-Eco RI	161	161	0
Eco RI-Taq I	128	129	3
	Hybridization Probe: NAR-1		
Eco RI-Sac I	101	101	0

except 7 sites appeared to be constant in both individuals. It is likely that this is a minimum estimate and that other cleavage site variants exist that are not detected in this analysis. Restriction fragments that have variant cleavage sites but co-migrate with other fragments or those that are not resolved by Southern blot hybridization under the conditions used would not be scored. The variation in the proportion of variant cleavage sites examined for each combination of enzymes and hybridization probe might reflect the degree of conservation of DNA sequence in the enzyme recognition sites flanking a particular repetitive DNA sequence.

DISCUSSION

The experiments presented here show that it is possible to use a variety of direct approaches to identify and characterize restriction fragment length polymorphisms in human DNA. Each of the three methods examined offers certain advantages. Chapter 1 demonstrates that random single-copy cloned DNA fragments isolated from recombinant DNA libraries can be successfully used to identify RFLPs. This approach can be used to generate polymorphic markers for arbitrary genetic loci, specific or nonspecific coding sequences of DNA, or for specific human chromosomes, depending on the types of recombinant DNA libraries used as sources of hybridization probes. Chapter 2 uses a fine structure mapping approach to identify differences in homologous DNA sequences which are reflected as RFLPs. This method is only limited by the number of different restriction endonucleases available. In Chapter 3, the use of middle repetitive DNA sequences as probes for restriction fragment length differences is explored. The preliminary studies described in this chapter examines a potentially useful technique that achieves resolution of individual restriction fragments hybridizing to middle repetitive DNA probes when analyzed by standard Southern blotting techniques. Recent evidence demonstrates the association of RFLPs with mobile repetitive elements in yeast and Drosophila DNA (31,32). Wyman and White (3) suggested that the molecular basis for a highly polymorphic locus in human DNA that was characterized by DNA rearrangements might be due to the presence of an active DNA element similar to those identified in yeast and Drosophila. No data exists for estimating the frequency of polymorphism in human DNA that might be due to this source. Insertions

and deletions of active DNA elements would be reflected in changes of restriction fragment lengths. If mobile DNA elements are a mechanism for generating restriction fragment length variation in humans, the methods employed in Chapter 3 might provide a useful tool for analyzing the contribution by these elements to DNA polymorphism.

Certain difficulties inherent in the present techniques of DNA hybridization and Southern blotting are common to all three methods employed to identify RFLPs. DNA fragments in the higher molecular weight ranges (>15 kb) are difficult to resolve, especially when they differ by as little as 0.5 kb. Very small DNA fragments, <0.2 kb, are not scored on most gels. Both sources of error act to lessen the sensitivity of the current methods employed to identify RFLPs. Very small DNA fragments represented in the genome as single-copy DNA sequences seldom hybridize effectively. This fact is illustrated by the inability to achieve consistent hybridization results when attempting to screen a panel of individuals for the two Hae III restriction fragment length differences described in Chapter 2.

Despite these limitations in the present technique, a number of restriction fragment length variants were identified. Two of these, an Eco RI cleavage site variant that hybridizes to clone G6, and a Taq I cleavage site variant that hybridizes to clone HD1A-H4, were shown to be polymorphic when hybridized to the DNAs from a panel of individuals comprising a Huntington's Disease pedigree. The allele frequencies of 0.56 for a restriction fragment lacking an Eco RI cleavage site when hybridized to clone G6, and 0.7 for a restriction fragment lacking a Taq I cleavage site when hybridized to clone HD1A-H4, were determined from a panel of 8 and 9 unrelated individuals, respectively. These estimates

must be regarded as very approximate until a larger sample is analyzed. A much more extensive screening would also be required to test for deviations from Hardy-Weinberg equilibrium.

Jeffreys (5) has used the data obtained from screening individuals for RFLPs with hybridization probes from the β-globin region of human chromosome 11 to estimate the overall degree of genetic diversity in the human genome. Three variant restriction sites were found out of a total of 52-54 different sites examined, representing a total of 300-310 bp of DNA sequence. The estimated frequency of polymorphism (0.01) was used to derive an estimate of $3X10^7$ polymorphic variants at the DNA level in the human haploid genome. Similar estimates can be derived from the data of this study. In Chapter 1, 14 hybridization probes were used to examine 133 restriction sites representing 798 bp of DNA sequence per individual. A total of 3 variant restriction sites were found. The frequency of polymorphism is calculated to be 3/798=0.0037. In Chapter 2, the analysis of two sets of matching clones surveyed a total of 185 restriction sites representing 866 bp of DNA sequence. Four variant restriction sites were found resulting in a frequency of 4/866=0.0046 for polymorphic variation. The data from Chapter 3 results in an estimation of polymorphic variation of 7/6400=0.0011 from the 7 restriction site variants scored out of an estimated 6400 bp of DNA sequence screened.

The following assumptions underlie the estimation of polymorphic variation in the human genome from the above data: A variant restriction site represents a single base change; DNA sequences in a restriction endonuclease site show the same level of genetic variation as noncleavable sequences; the degree of DNA sequence variation around

the area of DNA being examined by the hybridization probe is representative of the entire human genome. The estimated frequencies of polymorphism from the data of Chapter 1 (0.0037) and Chapter 2 (0.0046) are in close agreement with each other, but are less than the frequency (0.01) calculated from Jeffreys' data. The two sets of matching clones examined in Chapter 2 differ markedly with respect to the number of variant restriction sites. One set of clones was identical for every site examined. The other set of clones showed 4 restriction site variants for a total of 459 bp DNA sequence examined. A frequency of polymorphism calculated for this set of clones alone, 4/459=0.0087, is in closer agreement to the value estimated by Jeffreys. The 3 variant restriction sites in Jeffreys' data all occurred in intervening sequences leading to the suggestion that variants may cluster into highly variable "hot spots", perhaps in intervening sequences. It is interesting to note that 3 of the 4 variant restriction sites in the set of matching clones HD1A & HD1B are located in the same general area on either of the two adjacent 3.7 kb or 1.9 kb subfragments. The 4th variant site, an Alu I cleavage site, has not yet been localized. The variation in the estimation of polymorphism by each of the three methods used might reflect differences inherent in the sampling techniques, or differences in the degree of conservation of DNA sequences in the restriction sites examined. It is likely that these first data for polymorphic variation at the level of DNA sequence are not necessarily representative of the entire human genome.

The estimation of genetic variation at the level of DNA sequence by using restriction endonucleases in methods presented here might prove to be a useful method for studying genetic differences in regions of the

genome not accessible in other ways. One source of error that might lead to an overestimation of genetic variation when using this data is the variation in restriction fragment lengths caused by insertions or deletions of DNA sequences. Since only base substitutions are measured by these techniques, fragment lengths that are sufficiently altered by insertions or deletions will have homologous sites that will appear to be nonhomologous. This would result in an overestimation of the number of polymorphic sites.

A linkage map of the human genome can be constructed by determining the pattern of cosegregation of genetic markers defined by RFLPs in human pedigrees. Several hundred markers appropriately distributed should allow the mapping of any other locus with reasonable certainty (33). Human pedigrees (such as the Huntington's Disease pedigree which was used in this study) in which inherited disease traits are known to be segregating, can be tested for linkage of genetic disease loci to loci defined by RFLP markers. The development of markers that allow the mapping of genes responsible for inherited diseases will have an impact on predictive genetic counseling, antenatal diagnosis of genetic disease, and on resolving human genetic models of inheritance that are currently refractory to simple genetic analysis.

REFERENCES

1. Harris, H. & Hopkinson, D. A. (1972) Ann. Human Genet. 36:9-19.

2. Schafer, M. & White, R. 6th International Workshop on Human Gene Mapping, Oslo, Norway, 1981.

3. Wyman, A., & White, R. (1980) Proc. Natl. Acad. Sci. USA 77:6754-6758.

4. Kan, Y. W. & Dozy, A. M. (1978) Proc. Natl. Acad. Sci. USA 75:5631-5635.

5. Jeffreys, A. J. (1979) Cell 18:1-10.

6. Tuan, D., Biro, P. A., de Riel, J. K., Lazarus, H. & Forget, B. G. (1979) Nucleic Acids Res. 6:2519-2544.

7. Greever, R. F., Wilson, L. B., Ferez, S. N., Milner, P. F., Bittner, M. & Wilson, J. T. (1981) Proc. Natl. Acad. Sci. USA 78:5081-5085.

8. Higgs, D. R., Goodbourn, S. E. Y., Wainscoat, J. S., Clegg, J. B. & Weatherall, D. J. (1981) Nucleic Acids Research 9:4213-4224.

9. Beutler, E., Kuhl, W., & Johnson, C. (1981) Proc. Natl. Acad. Sci. USA 78:7056-7058.

10. Fiddes, J. C. & Goodman H. M. (1981) J. Mol. Appl. Genetics 1:3-18.

11. Rotwein, P., Chyn, R., Chirgwin, J., Cordell, B., Goodman, H. M., Permutt, M. A. (1981) Science 213:1117-1120.

12. Matthyssens, G. & Rabbitts, T. H. (1980) Proc. Natl. Acad. Sci. USA 77:6561-6565.

13. Hieter, P. A., Korsmeyer S. J., Waldmann, T. A. & Leder P. (1981) Nature (London) 290:368-372.

14. Brown, W. M. (1980) Proc. Natl. Acad. Sci. USA 77:3605-3609.

15. Giles, R. E., Blanc, H., Cann, H. M. & Douglas W. C. (1980) Proc. Natl. Acad. Sci. USA 77:6715-6719.

16. Hutchinson, C. III, Newbold, J., Potter, S., Edgell, M. (1974) Nature (London) 251:536.

17. Potter, S., Newbold, J., Hutchinson, C. III, Edgell, M. (1975) Proc. Natl. Acad. Sci. USA 72:4496-4500.

18. Avise, J. C., Lansman, R. A. & Shade, R. O. (1979) Genetics 92:279295.

19. Avise, J. C., Giblin-Davidson, C., Laerm, J., Patton, J. C. & Lansman, R. A. (1979) Proc. Natl. Acad. Sci. USA 76:6694-6698.

20. Ferris, S. D., Brown, W. M., Davidson, W. S. & Wilson, A. C. (1981) Proc. Natl. Acad. Sci. USA 78:6319-6323.

21. Southern, E. (1975) J. Mol. Biol. 98:503

22. Lawn, R. W., Fritsch, E. F., Parker, R. C., Blake, G. & Maniatis, T. (1978) Cell 15:1157-1174.

23. Kernitt, D., Kunkel, L. & Villa-Komaroff (1981) Am. Soc. Human Genetics, Abstract.

24. Benton, W., Davis, R. (1977) Science 196:800-802.

25. Rigby, P. W., Diekman, M., Rhodes, C. & Berg, P. (1977) J. Mol. Biol. 113:237.

26. Gusella, J., Varsanyi-Breiner, A., Kao, F., Jones, C., Puck, T. T., Keys, C., Orkin, S. & Housman, D. (1979) Proc. Natl. Acad. Sci. USA 76:5239-5243.

27. Schmid, C. & Deininger, P. (1975) Cell 6:345-358.

28. Blattner, R., Williams, B., Blechl, A., Denniston-Thompson, K., Faber, L., Grunwald, D., Kiefer, D., Moore, D., Schuman, J., Sheldon, E. & Smithies, O. (1977) Science 196:161-169.

29. Becker, A. & Gold, M. (1977) Proc. Nat. Acad. Sci. USA 72:581-585.

30. Duncan, C. H., Biro, P. A., Prabhakara, V. C., Elder, J. T., Wang, R. C., Forget, B. G., De Riel, J. K., & Weissman, S. M. (1979) Proc. Natl. Acad. Sci. USA 76:5095-5099.

31. Cameron, J. R., Loh, E. Y. & Davis, R. W. (1979) Cell 16:739-751.

32. Rubin, G. M. (1978) Cold Spring Harbor Symp. Quant. Biol. 42:1041-1046.

33. Botstein, D., White, W. L., Skolnick, M. & Davis, R. W. (1980) Am. J. Hum. Genetics 32:314-331.