THE DISCOVERY AND CHARACTERIZATION OF RESTRICTION FRAGMENT LENGTH POLYMORPHISMS IN THE HUMAN GENOME USING COSMID LIBRARIES

by

Linda Lee Bufton

A DISSERTATION

Presented to the Department of Medical Genetics
and the Oregon Health Sciences University
School of Medicine
in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy

April 1986

APPROVED:

[redacted]

(Professor in Charge of Thesis)

[redacted]

(Chairman, Graduate Council)

ACKNOWLEDGEMENTS

I am especially grateful to Dr. Michael Litt for his patient guidance, encouragement, enthusiasm, and full time availability.

I am also grateful for a pleasant working environment thanks to my coworkers Cheryl Maslen-McClure (who shared in each success and failure), Ruth Litt (always a ready listener with sensible advice), and Norm Buroker.

I thank my thesis committee members Drs. Ellen Magenis, Everett Lovrien, and Robert Koler for their advice and constructive criticisms. Special thanks to Dr. Lovrien for giving me my start in research in 1979, to Dr. Koler and the Department of Medical Genetics for supporting my pursuit of a combined M.D./Ph.D. program, and to Dr. Richard Jones and the Department of Biochemistry for treating me as one of their own.

Bob Sheehy and Sue Ellen Toth-Fejel kindly provided time, materials, and advice for preparation of figures and photographs. Rick Bestwick provided invaluable help with ligation and packaging methodology.

The financial burden of the last 2 years was eased considerably by a generous scholarship from the Life and Health Insurance Medical Research Fund.

CONTENTS

Page

TABLES

FIGURES

Page

SECTION II

Page

SECTION III

Page

ABBREVIATIONS

A	adenine
AFP	alpha-fetoprotein
APRT	adenine phosphoribosyl transferase
bp	basepairs
BCL	blood cell lysis
BSA	bovine serum albumin
C	cytosine
CDTA	diaminocyclohexane-tetraacetic acid
CEPH	Human Polymorphism Study Center
cM	centimorgans
cos	cosehive end site of λ phage
CsCl	cesium chloride
DNA	deoxyribonucleic acid
dpm	disintegrations per minute
dNTP	dexynucleotide triphosphates
EDTA	ethylenedinitrilo-tetraacetic acid, disodium salt
G	guanine
g	grams
HD	Huntington's disease
HGM	Human Gene Mapping workshop
HLA	human leukocyte antigens
HPRT	hypoxanthine-guanine phosphororibosyl transferase

kb	kilobases
L	liter
LMT	low-melting-temperature
M	molar
mCi	millicurie
mg	milligram
ml	milliliter
mM	millimolar
MW	molecular weight
µCi	microcurie
µg	microgram
µl	microliter
ng	nanogram
NL	nuclear lysis
OLB	oligolabeling buffer
p	phosphate group
PIC	polymorphism information content
PKU	phenylketonuria
RFLP	restriction fragment length polymorphism
RNA	ribonucleic acid
SDS	sodium dodecyl sulphate
SRO	shortest region of overlap
ss	size standard
T	thymine

ABSTRACT

The goal of this study was to discover restriction fragment length polymorphisms (RFLPs) in human DNA which have high degrees of heterozygosity and which should therefore be extremely useful markers for human gene mapping.

It has been estimated that 165 evenly spaced highly polymorphic markers throughout the human genome should allow the mapping of any gene or new marker with a high degree of reliability (Botstein et al, 1980; Lange & Boehnke, 1982). In order to obtain 165 evenly spaced markers, however, a considerably larger number of randomly spaced markers will be required. Of the 333 RFLPs reported in August 1985 at the Human Gene Mapping workshop in Helsinki, less than 10% have high heterozygosity, and their chromosomal distribution is very uneven (Willard et al, 1985).

This study used two strategies to reduce the number of random RFLPs needed to get 165 evenly spaced RFLPs spanning the human genome.

The first approach was to conduct the RFLP search using DNA libraries constructed in cosmids instead of the more often used phage libraries. Cosmids have larger human DNA inserts increasing the amount of DNA examined per clone and increasing the likelihood of finding multiple, closely linked loci. Such loci can be treated as a single compound locus which will most often have a higher heterozygosity than a simple locus. For example, a single 2-allele

locus could have a maximum heterozygosity of 50% whereas 3 closely linked 2-allele loci could yield up to 8 haplotypes with a maximum heterozygosity of 87.5%.

The second strategy was to use chromosome specific libraries which allowed the targeting of poorly mapped chromosomes for a more systematic RFLP search.

My initial work focused on a random cosmid clone from a total human DNA cosmid library. From this cosmid I discovered two subclones that revealed four closely linked RFLPs with a minimum heterozygosity of 72%. Using a somatic cell hybrid mapping panel and in situ hybridization to metaphase chromosomes, this compound locus was mapped to chromosome 19p13.2->19cen.

The bulk of my thesis research involved the construction and screening of a partial DNA library specific for human DNA on the long arms of chromosomes 11 and 16. These chromosomes were chosen because of the availability of a mouse-human somatic cell hybrid containing an 11q/16q translocation chromosome as its only human DNA and because very few RFLPs have been mapped to these parts of the human genome.

The partial screening of this library resulted in the discovery of two cosmids each of which reveals a highly polymorphic locus on chromosome 16q. Mapping to this chromosome was done by two methods, the use of a somatic cell hybrid mapping panel and in situ hybridization to metaphase chromosomes.

I. INTRODUCTION

Restriction fragment length polymorphisms, or RFLPs, represent a new set of genetic markers based on direct detection of deoxyribonucleic acid (DNA) sequence polymorphisms using restriction endonucleases (Botstein et al, 1980). RFLPs are being used to define a large number of arbitrary marker loci in man as well as some with known gene specificity. Using this new system, the construction of a detailed linkage map of the human genome is underway (White et al, 1985).

RFLPs are defined by cloned DNA segments which can be localized to specific regions of specific chromosomes and ordered on the chromosome using family linkage studies. It has been estimated that 165 evenly spaced highly polymorphic markers 20 centimorgans (cM) apart throughout the human genome should allow the mapping of any gene or new marker with a high degree of reliability (Botstein et al, 1980; Lange & Boehnke, 1982). An effective strategy for constructing a linkage map is to use a reference panel of normal, highly informative families, that is, families with a large number of children, both parents, and all four grandparents available for typing. Selection and analysis of markers for linkage in families exhibiting specific disease phenotypes will be most effective after such a map has been constructed (Botstein et al, 1980).

Prior to 1970, gene mapping primarily involved following the inheritance of traits, diseases, and observable protein polymorphisms

in families using classical linkage methods. In this classical sense, mapping a gene means determining its position in recombination units relative to other genes on the map; distances between loci depend on the extent of physical separation and on the rate of occurrence of crossing over. Except for genes assigned to the X chromosome based on X-linked recessive inheritance, most linkage groups were established without knowledge of chromosomal location. Stable variants in chromosome morphology were used as early as 1968 to assign some genes to specific autosomes. Donahue et al made the first such assignment when they demonstrated linkage between the Duffy blood group locus and the secondary constriction region of chromosome 1, 1qh (Ott, 1985). Similarly, haptoglobin was mapped to chromosome 16 using the fragile site on 16q (Magenis et al, 1970). With the advent of modern molecular biology, the scope of gene mapping has expanded greatly so that not only chromosomal assignment but also regional localization within the chromosome are common components of gene mapping.

Recent advances in molecular biology that have lead to current recombinant DNA techniques and the discovery of RFLPs include the discovery and characterization of restriction endonucleases, the development of DNA cloning strategies using plasmid and phage vectors and bacterial hosts, and improved electrophoretic and nucleic acid hybridization techniques.

A. Recombinant DNA technology

Restriction enzymes have been extensively used for studying DNA structure since Smith and colleagues showed in 1970 that an endonuclease from *Hemophilus influenza* cleaved double stranded DNA at a specific sequence (reviewed in Old & Primrose, 1985). Their discovery, coupled with improvements in transformation of *E. coli*, led to the development of DNA cloning strategies using plasmid and phage vectors. Agarose gel electrophoresis allowed high resolution separation of restriction fragments, an important step in preparing cloning vectors and in analyzing results of cloning experiments. Nucleic acid hybridization techniques such as those developed by Southern (1975) provided simplified ways of finding and visualizing homologous DNA fragments in complex genomes.

1. Restriction enzymes. Perhaps the first major discovery responsible for the development of recombinant DNA techniques was that of restriction endonucleases, endodeoxyribonucleases from bacteria that recognize specific nucleotide sequences in double stranded DNA and cleave both strands of the duplex. They were discovered nearly 35 years ago by Luria and Human, and the first one was characterized in 1968 (Nathan & Smith, 1975). The initially described enzymes, however, were of the Class I type; while they recognize specific nucleotide sequences, they do not cleave DNA at specific sites. It was the discovery of the Class II, or cleavage-site specific, endonucleases in 1970 by Smith and colleagues that lead to

the application of these restriction enzymes in recombinant DNA work (Nathan & Smith, 1975).

Restriction enzymes *in vivo* form a self protection system in bacteria against foreign DNA such as viruses; they represent a prokaryotic "immune system". The bacterial DNA is modified at the recognition site by methylation to render it resistant to restriction, thus protecting against cleavage of host DNA. Most Class II restriction endonucleases recognize and cleave DNA within a particular 4- or 6-nucleotide sequence. For example, EcoRI cuts the sequence GAATTC between the GpA on each strand. Many restriction enzymes make similar staggered cuts, resulting in protruding 5' termini. These termini are then available to associate by hydrogen bonding with similar termini, hence the designation of sticky or cohesive ends. Other Class II enzymes make cuts resulting in 3' cohesive ends while some produce blunt rather than staggered ends so that they are not cohesive.

Class II restriction enzymes have been isolated from a large number of bacteria, and they are still being discovered. About 500 have been at least partially characterized to date, but only about 1/10 of these have been well characterized and are used routinely by molecular biologists (Old & Primrose, 1985). With this wide range of currently available enzymes, cloning vectors and DNA inserts can be tailored in endless ways to achieve the desired results in DNA cloning experiments.

Restriction enzymes "establish convenient, fixed landmarks along the otherwise featureless terrain of the DNA molecule" (Weinberg, 1985). When a homogeneous DNA preparation is completely digested with a given Class II restriction enzyme, a characteristic set of fragments is produced. With DNA of low complexity, such as plasmids or λ phage, the restriction fragments can be separated by gel electrophoresis to give a fragment pattern that characterizes the original DNA molecule in much the same way that tryptic fingerprints characterize particular proteins. With more complex genomes, such as the human genome, one sees only a smear on the gel after digestion and electrophoresis. A locus of interest, arbitrary or specific, can be visualized by hybridization with radioactive single-copy cloned DNA segments or "probes". DNA sequence variation is detected by alteration in the length of one or more restriction fragments.

2. DNA libraries. A DNA library is a set of cloned fragments that represent the DNA from a specified source, e.g. a human genomic library contains a collection of cloned random DNA fragments representative of the entire genomic DNA of the organism.

Construction of DNA libraries first became feasible in the mid 1970s with the development of DNA cloning techniques. DNA cloning involves the amplification of DNA fragments from any source by inserting them into a plasmid or bacteriophage vector and growing these recombinant DNA molecules in bacterial host cells.

The first genomic DNA libraries were constructed for organisms with small genomes such as *Drosophila* (Wensink et al, 1974; Carbon et al, 1977) and yeast (Carbon et al, 1977). Construction of genomic DNA libraries for more complex genomes such as mammals became possible a few years later when several technical limitations were overcome including more rapid screening techniques for large numbers of clones (Benton & Davis, 1977; Hanahan & Meselson, 1980) and more efficient methods for introducing foreign DNA into host cells (Maniatis et al, 1978).

The components of a DNA library include: (1) the DNA of interest which has been isolated, purified and digested with an appropriate restriction enzyme, (2) the vector DNA which may be a plasmid, phage, or a cosmid, and (3) the host cell which is usually a strain of *E. coli*.

There are a variety of methods employed in the construction of DNA libraries. Based on the choice of vectors, 3 types are possible: plasmid, phage, and cosmid. The important difference is the maximum size of the foreign DNA insert that can be ligated to the vector which is about 15, 25, and 45 kilobases (kb) respectively.

Cosmids were first developed by Collins and Hohn (1978) specifically for cloning large fragments of DNA. Cosmids are synthetic hybrids of plasmids and phage with the following essential components: (1) a plasmid origin of replication, (2) one or more drug resistance genes, (3) one or more unique restriction sites for

cloning, (4) the cohesive end (cos) site of λ phage, and (5) a small size of 4-6 kb to accomodate large inserts (Maniatis et al, 1982).

Cosmid libraries were chosen for my RFLP searches because of the large insert size. The larger size increases the likelihood for finding multiple, closely linked RFLPs. The human haploid genome contains 3×10^9 base pairs (bp) of DNA or a total of 3300 centiMorgans (cM) in recombination units where one cM represents 1% recombination (Renwick, 1971) or about 10^6 base pairs (1000 kb). A cosmid with an insert of 40-45 kb would cover less than 0.05 cM so that multiple RFLPs revealed by a single cosmid would be very tightly linked at a recombination distance of <0.0005. Recombination between two such loci should be extremely rare.

High heterozygosity is critical to linkage studies, and the number and frequency of alleles characterizes the usefulness of a locus. In constructing genetic maps and calculating genetic risks when several linked markers are available, multilocus analysis is recognized as the most effective method (Lathrop et al, 1984, 1986). The success of such analysis depends on having single individuals that are heterozygous at all of the loci being examined. The majority of RFLPs reported to date represent 2-allele loci with low heterozygosity. Multiple loci revealed by single cosmids should be much more useful as genetic markers. For example, 3 closely linked loci could yield 8 haplotypes which, with equal frequencies, would give 87.5 % heterozygosity. While high degrees of linkage disequilibrium could diminish the usefulness of such loci,

observations to date indicate that such disequilibirium between closely linked RFLPs is often small (Litt & White, 1985).

The large insert size also allows the scanning of a larger piece of the human genome at one time with fewer colonies to screen. For example, to achieve a 99% probability that any single copy DNA will be represented in a cosmid library, one needs about 350,000 transformants versus about one million for a similarly complete phage library (Maniatis, 1982).

Cosmids have also been used to clone large genes on a single piece of DNA, for example, a 28 kb dihydrofolate reductase gene that could not be cloned in a single phage (Urlaub et al, 1985). Also, genes can be isolated along with large areas of their flanking sequences facilitating studies of gene regulation as has been done with the human interleukin 2 gene (Lindenmaier et al, 1985). Finally, chromosome walking can also be facilitated by the larger "steps" possible with cosmid libraries (Maniatis et al, 1982).

Ish-Horowicz and Burke (1981) reported one of the first uses of cosmids for eukaryotic libraries. They prepared a *Drosophila* DNA library using the cosmid vector pJB8 and the *E. Coli* host strain HB101. About the same time, Grosveld et al (1981) successfully constructed and screened a human DNA library using cosmids to isolate several recombinants containing inserts from the β globin-related genes. They also used the vector pJB8 and the host strain HB 101.

The presence of interpersed repeats in the human genome such as the Alu family initially posed a problem in using cosmids as probes to reveal polymorphisms. With the known occurrence of Alu sequences an average of once every 8 kb in the human genome (Lewin, 1985), nearly all cosmid inserts would be expected to contain at least one such sequence so that when the cosmids were used as probes the repeat sequences would hybridize at numerous genomic locations producing a smear on Southern blots that would obscure any single copy bands of interest. Litt & White (1985) have circumvented this problem by first prehybridizing the cosmid with a vast excess of unlabeled total sonicated human placental DNA under conditions that tie up the repeats so they are unavailable to hybridize on Southern blots. Using this method, cosmid probes have been used to reveal single copy RFLP bands without interference from genomic repeats (Litt & White, 1985; Litt et al, in press; Buroker et al, 1986; Bufton et al, 1986).

3. Gel electrophoresis and Southern blotting. The ability to efficiently follow and assess the DNA manipulations made possible with cloning techniques depends on rapid and simple techniques of DNA visualization. In the early 1970s, Sharp et al (1973) introduced agarose-ethidium bromide electrophoresis, which is used for both analysis and preparation of DNA. This method gives more precise results than the older method of velocity centrifugation through sucrose gradients and is much quicker and easier than sucrose gradients or acrylamide gel electrophoresis, the latter of which is

still used for isolating small DNA fragments of 100 bp or less. In additon to size separation, DNA molecules of different molecular configurations, e.g. closed circular vs. nicked circular or linear plasmids, can be separated due to different gel mobilities. Very small amounts of DNA, down to a few nanograms, can be detected by illuminating the gel with ultraviolet light since the DNA has been stained with the intercalating dye, ethidium bromide.

DNA mapping involves hybridization of DNA probes to restriction fragments of the DNA of interest. In 1975 Southern published a technique for transferring DNA fragments from a gel onto nitrocellulose membranes so that these membranes could be probed with radioactively labeled DNA and the hybridized fragments detected with autoradiography. This method, Southern blotting, proved far superior to previous methods that required cutting and eluting DNA from the gel and hybridizing the DNA in solution or after binding the eluted DNA to filters (Old & Primrose, 1985).

B. Restriction Fragment Length Polymorphisms

1. What are RFLPs and how are they detected? Differences among individuals in the length of a particular restriction fragment can result from point mutations that alter a restriction site or from rearrangements such as insertions and deletions that alter the distance between two restriction sites.

RFLPs are easily assayed in individuals using small amounts of DNA most often obtained from lymphocytes in peripheral blood

samples or from lymphoblast cell lines. RFLPs are inherited as simple Mendelian codominant markers so that inheritance and linkage relationships can be established using pedigree analysis (Botstein et al, 1980).

RFLPs can be used to map any trait caused wholly or in part by a major locus segregating in a pedigree without knowledge of the biochemical nature of the trait. These RFLPs need not be a part of the gene in question but will most often be closely linked, unrelated, random sequences. It has been suggested that RFLPs evenly spaced in the genome 20 cM apart should be adequate for mapping any Mendelian trait (Botstein et al, 1980; Lange & Boehnke, 1982).

At the Human Gene Mapping workshop (HGM 8) in August 1985 a total of 333 RFLPs had been reported in the human genome (Willard et al, 1985). However, their chromosomal distribution is very uneven and only about 10% are highly polymorphic with polymorphism information content (PIC) values >0.5. PIC is the probability that an offspring of a given mating will be informative, e.g. a PIC of 1 means that all offspring of a mating will be informative. PIC is calculated as:

$$1 - \left(\sum_{i=1}^{n} p_i^2\right) - \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} 2p_i^2 p_j^2$$

where n is the number of alleles and p_i and p_j are the allele frequencies (Botstein et al, 1980). Loci with many alleles and a PIC near 1 are most desirable. To achieve a goal of 165 evenly spaced, highly polymorphic markers, Lange & Boehnke (1982) estimate that

about 1500 randomly spaced markers will be needed if screening is from a total human genomic library but only about 750 if chromosome specific libraries are used.

2. How polymorphic is human DNA? The usefulness of DNA markers in gene mapping depends in part on the degree of polymorphism. A locus is considered polymorphic if it has a minor allele frequency of at least 10%. Early estimates of overall polymorphism were based on the frequency of protein polymorphism in man. Harris and Hopkinson (1972), summarizing the electrophoretic data from 71 enzyme loci, found 28% polymorphism which correlates with a DNA sequence heterozygosity of 0.28 bases per 1000 bases, the size of the average gene (Botstein et al, 1980), or 0.00028 per base pair. As visible electrophoretic changes represent only about 1/3 of amino acid changes and many third base changes do not cause any amino acid change (Harris, 1980) this would correspond to a DNA sequence polymorphism of about 0.001 per base pair among protein coding sequences. This figure represents a minimum value as sequences in messenger RNA diverge much more slowly than the bulk of single copy DNA (Rosbach et al, 1975).

A maximum figure for DNA polymorphism has been estimated at .011, or about 10 times the above figure, based on a single study of the depression of the melting temperature of reassociated DNA from 2 individuals of the same mammalian species (Botstein et al, 1980).

Using these minimum and maximum figures, Botstein et al (1980) calculated the possible number of RFLPs to be found per fragment with 4-base and 6-base recognition restriction enzymes at 0.0034-0.123 and 0.005-0.177 respectively. The 4-base cutters should be more useful, however, as they produce sixteen times more fragments per unit length of DNA (cutting every 4^4 vs 4^6 bp). Even the above figures may be conservative as they do not take into account polymorphisms due to rearrangements such as insertions and deletions, a number that cannot yet even be estimated.

Some studies have looked at the variation in genomic DNA at specific sites such as the β-globin region. Jeffreys (1979) estimated that one per hundred, or .01, base pairs in the β-globin gene region is polymorphic. These data were based on examining 41 kb of sequences over the ϒ, δ, and β globin loci. They used 8 different restriction enzymes to examine 53 restriction sites which represent a total of 310 base pairs of sequence. Three sites were polymorphic. If these polymorphisms represent single base changes as was thought, then 3 of 310 base pairs showed polymorphic variation for a rate of about 1 per 100 base pairs or 0.01. Only a small fraction of sequence variation would create or destroy a detectable restriction cleavage site. In the above study 41 kb of DNA was covered with restriction sites representing 310 base pairs, giving a variant detection rate of 0.7%. The use of more enzymes, however, would presumably increase the detection rate.

More recently Cooper et al (1985) estimated unique DNA sequence heterozygosity based on the number of RFLPs detected with 19 random cloned DNA probes used on total human DNA digested with 6 different enzymes. They came up with a DNA polymorphism frequency of 0.008 which agrees with Jeffreys (1979) and falls in the upper range postulated by Botstein et al (1980).

Based on results to date with various restriction enzymes, the selection of enzymes used in search of RFLPs is of importance. As previously mentioned, one would expect more RFLPs with 4-base cutters as more sites are present for these enzymes on the average than for 6-base cutters, sites every 256 vs every 4096 base pairs. While this general idea holds up, the actual fragment lengths and the number of polymorphisms is also related to the restriction site base pair sequence. Bishop et al (1983) developed a model for restriction fragment length distribution based on DNA dimer frequencies in humans by computing mean fragment lengths for known restriction enzymes. As predicted, 4-base cutters had shorter mean fragment lengths than 6-base cutters. In looking at the dimer frequency, however, they found a considerable deviation from expected values as had been previously noted (Swartz et al, 1962). CpG dimers are rare so that the enzymes with a recognition sequence containing this dimer have higher mean fragment lengths. For example, Taq I and Msp I (TCGA & CCGG) have mean fragment lengths of 1,179 and 1,747 base pairs, respectively, while several other 4-base cutters without CpG in the recognition sequence range from 198-493 base pairs per

fragment. The situation is similar for 6-base enzymes with CpG in their recognition sites. This is where the idea of "more sites/more polymorphisms" breaks down. Several studies (Barker et al, 1984; Cooper et al, 1985; Knowlton et al, 1985) have found a higher rate of polymorphism using Taq I and Msp I than with other 4-base cutters with smaller mean fragment lengths but without CpG in their recognition sequences.

Barker et al (1984) looked for RFLPs in humans using 32 single copy probes on DNA digested with five 4-base and eleven 6-base cutters. The two with CpG in their recognition sequences, Taq I and Msp I, revealed 9 of 10 total RFLPs found. From these results they estimated the frequency of polymorphism at sites containing CpG to be 1/10 to1/20 making them "hotspots" for mutation.

These findings are supported by additional data from both prokaryotes and eukaryotes. In bacteria, cytosine that deaminates to form uracil is efficiently corrected while methylated cytosine that spontaneouly deaminates to thymine is not efficiently corrected so that C to T transitions represent a hotspot for mutation (Lewin, 1985). As 2 to 7% of mammalian DNA is methylated at cytosine (Lewin, 1985), methylated CpG could be a frequent site of mutation in humans as well.

Savatier et al (1985) sequenced a 5.5 kb fragment of the β globin gene from chimpanzee DNA for comparison with the corresponding human sequence and found a much higher than expected divergence at CpG dinucleotides when compared with the overall very

low level of nucleotide sequence divergence between these two related species; they concluded that CpG instabilility exists.

At HGM 8 a group from Collaborative Research, Inc reported that in screening over 1500 random clones for RFLPs, they found the following enzymes most useful, listed in the order of efficiency: Msp I, Taq I, Rsa I, Pst I, BamHI, EcoRI, and Hind III (Knowlton et al, 1985).

3. How are RFLPs being used? RFLPS were first used as a tool for genetic analysis in 1974 when Grodzicker et al (1974) used the restriction maps of two serotypes of adenoviruses and their hybrid recombinants to map the physical location of temperature sensitive mutations. About the same time Hutchison et al (1974) used mitochondrial DNA restriction patterns to show maternal inheritance of this DNA in mammals.

Some of the first described human RFLPs were found in and around the globin genes (Lawn et al, 1980) which is not surprising as the globins represent one of the most extensively studied and characterized protein groups. Accordingly, with a good source of globin mRNAs in reticulocytes, they were one of the first and most studied higher eukaryotic gene families as recombinant DNA techniques became available. One of the first observed DNA polymorphisms in the human β globin gene was reported by Kan and Dozy (1978). They used a Hpa I restriction site polymorphism to diagnose sickle-cell anemia antenatally . Since then many more RFLPs have been found including several associated with

hemoglobinopathies such as sickle cell disease and the thalassemias. Antonarakis et al (1985) in a recent review list 17 "useful" polymorphic restriction sites in the β globin gene cluster which involve both coding and noncoding DNA.

These polymorphisms have proven useful in several ways including the carrier detection and prenatal diagnosis of sickle cell disease and the thalassemias, the elucidation of the evolution of the globin genes by studying the origin and spread of common mutations, and insight into normal gene structure and function and the nature of molecular defects that lead to disease (Antonarakis et al, 1985). Such intensive study of the globin genes also gives us good comparative data to use as other genes are similarly studied.

RFLPs are now being found in or around an ever increasing number of other known gene loci. A total of 88 polymorphic cloned DNA sequences associated with known genes were reported at HGM 8 (Willard et al, 1985), some of which include albumin, thyroglobulin, insulin, phenylalanine hydroxylase, alpha-1-antitrypsin, Factors VIII & IX, hypoxanthine-guanine phosphororibosyl transferase (HPRT), alpha-fetoprotein (AFP), human leukocyte antigens (HLA), and the immunoglobulins. Prenatal and carrier detection strategies are being developed for a number of diseases caused by defects in some of these genes, e.g. hemophilia (Graham et al, 1985; Oberle et al, 1985), α-1-antitrypsin deficiency (Cox et al, 1985), and phenylketonuria (PKU) (Woo et al, 1983). In addition, RFLPs are being used to study heterogeneity of the mutations that may lead to the

same disease, such as with Lesch-Nyhan syndrome (Yang et al, 1984).

The above mentioned RFLPs were recognized because of their relationship to a specific gene of interest, e.g. β globin, and were found subsequent to cloning of all or part of the gene. As previously stated, however, a more general use of RFLPs is the mapping of arbitrary loci which need not involve protein encoding DNA. One of the first such highly polymorphic arbitrary loci in human DNA was reported by Wyman and White in 1980. Using such RFLPs, genes of interest can be located without isolating them. The first such example using RFLPs was localization of the Huntington's disease (HD) gene to chromosome 4 by Gusella et al (1984). Using the knowledge that a RFLP closely linked to a gene would be inherited with that gene, they began tracing the inheritance in HD families of numerous RFLPs using random cloned DNA fragments from a human gene library as probes. Predicting that success would require looking at several hundred probes, they were incredibly lucky to find a closely linked RFLP in the first 12 cloned fragments screened. This work has lead to the development of presymptomatic diagnosis for this disease, and more importantly, to further work to bracket the gene with more markers and eventually isolate the gene itself.

Similar studies have lead to localization of the gene for cystic fibrosis on chromosome 7 (Knowlton et al, White et al, Wainwright et al, 1985) and the gene for adult polycystic disease on chromosome 16 (Reeder et al, 1985).

Perhaps the most powerful example of the use of RFLPs to date is the mapping of the X chromosome. The number of known RFLPs varies greatly between chromosomes with a minimum of 3 on chromosomes 10, 18, 20, & Y, and a maximum of 68 on the X chromosome (Willard et al, 1985). This is partly due to the disparity in the number of investigators working on any given chromosome and partly due to size differences of the chromosomes. There is still a question as to whether some chromosomes may be more polymorphic than others by virtue of their DNA content.

Long before development of recombinant DNA techniques, the X chromosome was the most studied chromosome because of its unique inheritance that allows identification of recessive X-linked traits. Using segregation analysis, more genes have been mapped to the X than to any other chromosome, e.g. many important genetic diseases such as Duchenne muscular dystrophy, fragile X syndrome, and Hemophilia A & B, have long been mapped to the X chromosome.

Now, by combining map information from previous studies with arbitrary loci defined by DNA RFLPs, an extensive linkage map of the X chromosome has been published (Drayna & White, 1985). A total of 21 markers were used, including 5 gene loci and 16 arbitrary loci, to span the entire distance of the chromosome, a total length of 185 recombination units. Such a map is useful in studying DNA behavior such as recombination and in comparison of genetic vs. physical distance. It is also useful in the study and diagnosis of X-linked disease. Duchenne muscular dystrophy has now been well bracketed

with RFLPs proving useful in carrier detection and prenatal diagnosis (Bakker et al, 1985) and two groups are now close to isolating the gene itself (Kolata, 1985; Kunkel et al, 1985; Monaco et al, 1985).

Polymorphisms resulting from hypervariable minisatellite regions in human DNA have such high heterozygosities that some are proving useful as DNA "fingerprints", DNA patterns completely specific to individuals (Jeffreys, 1985a, b, c). Minisatellites are short, simple tandem repeats dispersed in the human genome whose polymorphism results from allelic differences in the number of repeats. Such length polymorphism may arise by mitotic or meiotic unequal exchanges or by DNA slippage during replication (Jeffreys, 1985a). A probe representing the repeated core sequence can detect many highly variable loci simultaneously to provide an individual-specific DNA fingerprint. Such DNA fingerprints should prove useful in forensic science and paternity testing, and Jeffreys et al (1985c) report successful use of this method in an immigration case in which a mother questioned whether or not a child returned to her was really her son.

RFLPs are being used to follow the course of engraftment in bone marrow transplants (Blazar et al, 1985). Highly polymorphic RFLPs, such as those resulting from minisatellites, allow the discrimination of host vs donor cells when the donor may be very closely related to the recipient.

RFLPs are also proving useful in cancer research to determine clonal origin of tumors and to study the loss of heterozygosity at

various gene loci (Vogelstein et al, 1985; Dracopoli et al, 1985). Nonrandom deletions of chromosomal regions 13q14 and 11p13 have been detected in retinoblastoma (Cavenee et al, 1983) and Wilm's tumor (Koufos et al, 1984), respectively, using RFLPs.

One can see from the examples presented here that there is an explosion in information about the human genome with the advent of RFLP technology. Only 24 polymorphic cloned DNA segments were reported at HGM 6 in 1981 compared with 333 at HGM 8 in 1985 (Willard et al, 1985).

In addition to their proven use in gene mapping, RFLPs are also being used for gene isolation, carrier detection and antenatal diagnosis, studies of normal and abnormal gene structure, studies of evolution via recombination and other events, and studies of cancer-associated genes and cancer progression. They will probably prove useful in forensic science and paternity testing (Jeffreys et al, 1985c) and will also be useful in determining the genetic basis of multifactorial traits and disorders that cluster in families such as schizophrenia, epilepsy, and breast cancer (Botstein et al, 1980).

C. Thesis objectives

The overall thesis objective was to find and characterize several highly polymorphic RFLPs in the human genome and to establish the chromosomal location of these RFLPs.

For Part 1 a partial DNA library specific for human DNA on the long arms of chromosomes 11 and 16 was constructed in cosmids and

screened for highly polymorphic RFLPs on 16q. These chromosomes were chosen because of the availability of a mouse-human somatic cell hybrid containing as its only human DNA an 11q/16q translocation chromosome and also because very few RFLPs have been mapped to these parts of the human genome. At the most recent gene mapping workshop, 21 polymorphic loci had been assigned to chromosome 11, the majority on the short arm, and 6 RFLPs, all with low heterozygosity, assigned to chromosome 16 (Willard et al, 1985). 16q and 11q together represent about 5% of the human genome or about 165 cM (Daniel, 1985). Using Lange & Boehnke's data (1982), one would need to find at least 37 RFLPs from a chromosome specific library in order to have 3 to 5 evenly spaced RFLPs 20 cM apart on 16q and a similar number on 11q.

The plan was to construct a partial genomic DNA library using cosmid vectors and DNA from the somatic cell hybrid. A complete library was not necessary as I was looking for several random RFLPs on 16q rather than a specific singly copy sequence.

The library, once constructed, would be screened with total human and total mouse DNA probes to select those clones containing human inserts which should represent about 2.5% of the total clones. Clones with human inserts would be screened for those identifying RFLPs and the latter further screened and studied to isolate single copy probes useful for specific, highly polymorphic RFLPs. Gusella et al (1980) first showed that repetitive sequences in mammals had diverged sufficiently between species so that total human and total

rodent DNAs could be used as species-specific probes by using hybridization conditions under which only the repetitive sequences would give a detectable hybridization signal.

Chromosomal location was to be verified using a rodent-human hybrid mapping panel with more precise localization to be done by in situ hybridization in E. Magenis' laboratory.

The use of somatic cell hybrids to make chromosome specific libraries was first reported by Gusella et al (1980) and has been used for finding genes and RFLPs on several human chromosomes including 11 (Gusella et al, 1980) and 13 (Cavenee et al, 1984). All previous studies, however, have used phage rather than cosmid libraries. The technique of screening libraries with total rodent and total human DNA to pick clones containing human inserts has accordingly been reported to date only for phage libraries.

Part 2 is a manuscript from the April 1986 issue of The American Journal of Human Genetics. The research for this paper was actually done prior to most of the work for Part 1. It involved working with a random cosmid from the Ed Fritsch library which, on preliminary screening, had revealed several possible polymorphisms. The cosmid did indeed reveal several highly polymorphic RFLPs which were characterized and close linkage verified by family studies. The RFLP was mapped to chromosome19 using somatic cell hybrid panels provided by G. Bruns, D. Shaw, and D. Brook, and more precise mapping was done in E. Magenis' laboratory by in situ hybridization to metaphase chromosomes.

II

Part 1: The discovery and characterization of restriction fragment length polymorphisms on chromosome 16q using a cosmid library made from a mouse-human somatic cell hybrid

A. Materials and Methods

Appendix B lists the composition of all solutions.

1. Construction of the 11q/16q DNA library in cosmids.

a. Somatic cell hybrid DNA preparation. The somatic cell hybrid, CF 52-46-1/8, containing the translocated human chromosome 11q/16q, was from T. Mohandas, Harbor-UCLA Medical Center, Los Angeles, CA. The cells, received as frozen pellets, were diploid mouse cells containing as the only human DNA at least one translocated 11q/16q with about 15% of cells containing two or more copies of this chromosome (T. Mohandas, personal communication). Chromosome analysis was done in Dr. Mohandas' laboratory. As 11q + 16q represent about 5% of the haploid genome (Daniel, 1985), this translocated chromosome should represent at least 2.5% of the somatic cell hybrid DNA.

High molecular weight DNA (>200 kb) was isolated from the cells using procedures combined from Maniatis el at (1982) and Kunkel (Bell et al, 1981). About 1.2 x 10^8 cells were received frozen in 2.5 ml of 10mM Tris Cl^-, 150mM NaCl, 10mM EDTA, and 10% glycerol. They were thawed and pelleted at 2 Krpm at 4^0C for 5 minutes and the supernatant discarded. The cells were resuspended in 50 ml of BCL buffer, centrifuged at 2 Krpm at 4^0C for 20 minutes, the supernatant discarded and the pellet drained briefly. The pellet was gently resuspended in 12 ml ice cold NL buffer on ice. To another 12 ml of NL buffer was added 2.5 mg Proteinase K (Boehringer Mannheim) and 1.2 ml 10% SDS. While gently mixing this second solution, the

nuclear suspension was added to it dropwise. The nuclear lysate was incubated overnight at 37^{0}C with gentle rotation. The sample was then extracted twice with an equal volume of phenol/chloroform (1:1), and once with chloroform for 30 minutes with gentle rolling on a rotating platform, centrifuged at 2.8 Krpm for 10 minutes and the top layer removed to a new tube. After dialysis overnight at 4^0 C against 4 L of dialysis solution A, the sample was transferred to a 50 ml tube and treated for 2 hours at 37^{0}C with 100µg/ml RNAse. 100µg/ml Proteinase K and 0.5% SDS were added and the sample was incubated another hour at 37^{0}C. The sample was extracted twice with phenol/chloroform and once with chloroform using gentle rotation, dialyzed overnight at 4^{0}C against 4L of TE, loaded dropwise onto at least 2 volumes of 1M NaCl in TE in 38 ml polyallomer tubes and centrifuged overnight in a Beckman SW27 rotor at 25 Krpm. The supernatant was poured off and the tubes drained 10 minutes. The pellets were dissolved in a small volume of TE overnight at 4^{0}C with gentle shaking. The yield, checked by UV spectrophotometry, was about 500 µg per 10^8 cells (assuming that 1 mg/ml of DNA has an A_{260} = 20 in a 1 cm cuvette). The size was estimated at about 200 kb by electrophoresis in a 0.3% agarose gel using uncut λ DNA for comparison.

The high molecular weight DNA was partially digested with one unit of Sau 3A per 128 µg DNA using the conditions determined by a series of test digests to give a maximum number of fragments in the 35-45 kb size range (Maniatis et al, 1982). The enzyme was heat

inactivated by incubating the digest at 68^{0}C for 10 minutes, and the DNA was ethanol precipitated and dissolved in TE. The partial digest was then size fractionated using a 10-30% sucrose density gradient with 10% and 30% sucrose solutions made in 1M NaCl, 20mM Tris Cl^-, and 5mM EDTA (Maniatis et al, 1982). Samples were layered onto the gradient in Beckman 12.8 ml ultra-clear centrifuge tubes and were centrifuged in a Beckman Sw41 rotor at 23Krpm for 14 hours at 20^{0}C. Fractions were collected using a rigid plastic tube connected to flexible tubing which was connected to a Gilson minipuls pump. The plastic tube was placed vertically into the sample tube so that it was almost touching the bottom, and 0.75 ml fractions were collected using this pump set-up with a Gilson fraction collector. Aliquots were analyzed on a 0.4% agarose gel and those containing DNA in the 35-45 kb size range were pooled, ethanol precipitated, and the pellets dissolved in TE.

b. Vector DNA. The cosmid vector pJB8, shown in figure 1, was used for this library. Prepared pJB8 arms with end BamHI sites (steps 1-3, figure 1) were obtained from Amersham Corporation, Arlington Heights, IL.

c. Ligation, *in vitro* packaging, and transduction into host bacteria. The general scheme used is shown in figure 1 (modified from Maniatis et al, 1982). Ligations were done at a total DNA concentration of 165 µg/ml with the molar ratio of vector molecules to insert at 1:1:1 as the desired concatamer was vector arm 1 (HindIII/Bam HI) : insert : vector arm 2 (BamHI/SalI) (Maniatis et al,

1982). The DNA was ligated overnight at 15^{0}C in 50mM Tris Cl^-, pH7.4, 10mM $MgCl_2$, 1mM spermidine, 5mM dithiothreitol, 1mM ATP, and 100 µg/ml bovine serum albumin with 1 unit of Boehringer Mannheim T_4 DNA ligase per 20 µl reaction for 3 µg of DNA. Ligation was monitored on a 0.3% agarose gel by disappearance of unligated vector arms and increased size of the smear of insert DNA due to concatamerization.

Note in figure 1 that after ligation an entire complement of plasmid sequence is contained between the 2 *cos* sites. Concatamers are efficiently packaged if the *cos* sites, the substrates for a packaging dependent cleavage, are 37-52 kb apart or 75-105% the size of λ DNA (Old & Primrose, 1985). Following *in vitro* packaging and introduction into *E. coli*, the cosmid DNA recircularizes and replicates in the form of a large plasmid which contains the β-lactamase gene that confers ampicillin resistance on the host bacteria (Maniatis et al, 1982).

The ligated DNA was packaged into bacteriophage λ particles *in vitro* using the "Packagene" extract system from Promega Biotec according to the manufacturer's instructions, with 0.5 µg of DNA and 1/2 tube of extract per reaction. The packaged DNA was transduced into 3 different *E. coli* host strains: HB101, ED8767, and DK-1. For use in transduction, the bacteria were grown up overnight in L broth with 0.4% maltose, pelleted, and resuspended in 0.5 volume of 10mM $MgSO_4$ or 10mM $MgCl_2$. Fifty microliters of packaging reaction was

mixed with 100 µl SM and 200 µl of host bacteria. The λ particles were allowed to adsorb at 37^{0}C for 20 minutes, then 1 ml of L broth was added and the mixture was incubated at 37^{0}C for 45 minutes to allow expression of ampicillin resistance. The cells were pelleted, resuspended in 200 µl of L broth and spread on day old L agar + 200 µg/ml ampicillin plates. Plates were inverted and incubated overnight at 37^{0}C.

Two methods were used for storing parts of the library. The first method was transfer of transformed colonies from the original plates into 150 µl of freezing medium with 200 µg/ml ampicillin in 96-well tissue culture plates which were kept at -70^{0}C (Schleif & Wensink, 1981). The second method was to toothpick transformed colonies from the original plates onto 150x15mm L agar + 200 µg/ml ampicillin plates using a grid template for 520 colonies per plate. Plates were wrapped and stored at 4^{0}C.

2. Screening the cosmid library.

a. Preparation of replica filters. Replica filters were made from the master plates and processed using procedures modified from Grunstein & Hogness (1975). Colonies were transferred from the 96-well plates to sterile nitrocellulose filters (Millipore HATF or Schleicher & Schuell BA 85) on L-agar + 200 µg/ml ampicillin plates using a stainless steel 96-pin replicator (West Coast Scientific). Replica filters were made from the master grid plates by placing a sterile filter on the surface of the toothpicked plate, notching the filter with an 18 gauge needle at asymmetric points

also marked on the master plate, then transferring the filter, colony side up, to a fresh L-agar + ampicillin plate. In both cases the filters were incubated at 37^{0}C until colonies were visible.

Replica filters were prepared for colony hybridization as follows: they were placed, colony side up for 5 minutes each, on a series of 3 3MM Whatman filter papers moistened with 10% SDS, 0.5M NaOH/1.5M NaCl, and 1.5M NaCl/0.5M Tris Cl^-, pH7.5, respectively, with brief blotting on paper towels between steps. They were then wetted with 1X SET, dipped through a 100µg/ml solution of Proteinase K in SET and placed on SET-moistened 3MM sheets to incubate for 30 minutes at room temperature. They were then dipped through a solution of SET, dried on paper towels, sandwiched individually between 3 MM sheets, and baked in a vacuum oven at 60-80^{0}C for 1-2 hours. The filters were then washed at 45^{0}C in prewash solution for 1-2 hours. Finally they were prehybridized in seal-a-meal bags overnight at 45^{0}C in prehybridization/hybridization solution (without dextran sulfate).

F filters, those stored at -70^{0}C as back ups, were made and incubated on L-agar + ampicillin plates containing 5% glycerol. They were prepared for freezing by placing them on sterile 3MM filter paper, placing a second freshly wetted filter on top, sandwiching the two filters together between more 3 MM sheets by pressing with a heavy weight, and placing the sandwich with a moistened sheet in a seal-a-meal bag (Hanahan & Meselson, 1980).

At one point, three of the F filters were thawed, several cosmid-containing HB 101 clones toothpicked from each filter into L-Broth, and fresh overnight cultures grown up. Intact cosmid DNA purified from these overnight cultures was then packaged *in vitro* and transduced into DK-1.

b. Colony screening. To determine which clones contained human DNA inserts, one set of replica filters was probed with radioactively labeled total mouse DNA and a duplicate set with total human DNA. Total mouse DNA was prepared from mouse fibroblast line GM 346-A9 provided by T. Mohandas using the high molecular DNA preparative method previously described. Total human DNA from a random individual was prepared from white blood cells as described in section 3a. The mouse and human DNAs were labeled using the "oligo" or "primer extension" method of Feinberg and Vogelstein (1983) referred to subsequently as oligolabeling. The DNA was denatured at 100^0C for 5 minutes then incubated with a mixture of hexanucleotide primers and deoxynucleoside triphosphates (dNTPs), including α-^{32}P deoxycytidine triphosphate (dCTP) or α-^{32}P deoxyadenosine triphosphate (dATP). Radioactive α-^{32}P dCTP and dATP with a specific activity of 10 millicuries(mCi)/ml were from New England Nuclear.

A typical reaction was done in 25 µl consisting of 10 µl OLB, 2 µl 1% gelatin, 5 µl α-^{32}P dATP (50 µCi), 2 units Klenow Polymerase I (Boehringer-Mannheim or Pharmacia), and 8 µl DNA (64 ng). Incubation was at room temperature for at least 2.5 hours, usually

overnight. The percent incorporation was determined by spotting 0.1-0.5 µl of the reaction mix on Whatman GF/C glass fiber filter paper and scintillation counting before and after washing the filters with 5% trichloroacetic acid/0.1% sodium pyrophosphate. Specific activities ranged from 3×10^8 - 1×10^9 dpm/µg DNA.

The colony filters were hybridized in 1 ml of hybridization solution (same as prehybridization solution) per filter with 4×10^6 dpm/ml of radioactive total mouse or total human DNA in seal-a-meal bags overnight at 44-45^{0}C. The filters were washed once in 2X SSC/0.1% SDS at room temperature for 15 minutes, twice in 2X SSC/0.1% SDS at 55^{0}C for 30 minutes and once in 1X SSC/0.1% SDS at 55^{0}C for 30 minutes. They were dried on paper towels, wrapped in Saran wrap, taped to cardboard backing containing ^{14}C India ink orientation marks, and exposed overnight to Kodak XAR-5 film backed by a Dupont Cronex Lightning-Plus Intensifier at -70^{0}C.

c. Preparation of DNA from cosmids containing human inserts.

The rapid alkaline extraction method of Birnboim (1983) was used with some modifications. A fresh 5 ml overnight culture of the clone of interest was grown up in L broth + 200 µg/ml ampicillin by inoculation from the master plate. One and a half milliliters of the culture was pelleted in an Eppendorf tube by a 15 second centrifugation, the supernatant discarded, and the pellet loosened by vortex mixer. The pellet was suspended in 200 µl lysozyme solution with 1 mg/ml fresh lysozyme (Sigma) and iced for 5 minutes. 400 µl alkaline SDS was added, the solution mixed gently by inversion

3-4 times, placed on ice 5 minutes, 300 µl high salt solution added, the solution again mixed gently by inversion and iced 15 minutes, and the solution centrifuged at room temperature for 2 minutes. 700 µl of the supernatant was transferred to 2 fresh tubes and the DNA precipitated with 2.5 volumes of ethanol, pelleted and dissolved in acetate-MOPS, reprecipitated and again dissolved in acetate-MOPS, reprecipitated and dissolved in 80 µl TE^{-4}. DNA size and quantity were estimated by electrophoresis in a 0.4% agarose gel with known cosmids as size standards.

Larger scale cosmid DNA preparations were also attempted using similar alkaline lysis followed by centrifugation in cesium chloride-ethidium bromide density gradients (Maniatis et al, 1982).

d. Cosmid subcloning. Cosmids revealing possible polymorphisms were subcloned into the plasmid pSP65 (Promega Biotec), a 3 kb ampicillin resistant plasmid with a polylinker containing 11 unique restriction sites for cloning. Cosmid DNA was digested with 10X Sau 3A, phenol/chloroform extracted, ethanol precipitated, dissolved in TE^{-4}, and ligated into the BamHI site of Bam HI cut and phosphatased pSP65 using 100ng of insert DNA per 1 µg of vector DNA with reagents and conditions as previously described. Competent HB101 bacteria were transformed with the recombinant plasmids using standard techniques (Maniatis et al, 1982) and transformed colonies selected on ampicillin plates. Master plates and replica filters were made as previously described for cosmids except that 85mm diameter filters were used with 100

colonies per filter. The clones were screened by colony hybridization using oligo-labeled total human DNA and appropriate oligolabeled restriction fragments from the cosmid. DNA was isolated from the subclones of interest by the Birnboim alkaline miniprep method (Birnboim, 1983).

3. Southern blotting and hybridization.

a. Source and preparation of human DNA. Human DNAs were prepared from outdated whole human blood obtained from the local blood bank and from lymphoblast cell lines of large Utah families obtained from the Human Genetic Mutant Cell Repository, Institute for Medical Research, Camden, NJ, using the method of Kunkel (Bell et al, 1981) with the addition of a second ethanol precipitation in the presence of 2.5M ammonium acetate. Restriction enzymes were obtained from Promega Biotec, New England Biolabs, Bethesda Research Labs, and Boehringer Mannheim and used according to the manufacturers' instructions. Human DNAs were digested with 5-10 units of enzyme per microgram of DNA and completeness of digestion assessed by agarose gel electrophoresis of parallel digests containing λ DNA in addition to human DNA (Barker et al, 1984a). Complete digests were ethanol precipated and dissolved in TE.

b. Southern blotting. The digested total human DNAs were electrophoresed through agarose gels, transferred to nylon membranes (Gene Screen from NEN, Genatran from D&L Filter, Woburn, MA, or Zetapore from AMF-CUNO, Meriden, CT), washed and prehybridized according to Barker et al (1984b).

c. Hybridization. Radiolabeled whole cosmids were screened on the Southern blots for their utility in revealing RFLPs. They were nick-translated (Barker et al, 1984a) in the presence of α-^{32}P dATP to give specific activities of at least 2 x 10^8 dpm/µg. After removal of unincorporated radioactivity by spermine precipitation (Hoopes & McClure, 1981), the cosmid probes were mixed with a large excess (625 µg/100-200 ng cosmid DNA) of non-radioactive 2.5 mg/ml sonicated (500 bp) human placental DNA (Calbiochem), heated at 100^{0}C for 10 minutes, and prehybridized to a c_0t of about 100 moles-sec/L by incubation in 0.12M Sodium phosphate, pH 7, at 65^0 C for 4-6 hours (Litt & White, 1985).

These prehybridized probes were then hybridized with Southern blots of Taq I and Msp I digested genomic DNAs from a panel of unrelated individuals. One lane containing the somatic cell hybrid line CF 52-46-1/8 DNA was also included on later blots to detect cosmids containing mouse DNA that were missed in the first screening. Hybridization was overnight at 43-45^{0}C in hybridization solution with dextran sulfate. The blots were then washed in 2X SSC/0.1% SDS at room temperature for 15 minutes, once in 0.1X SSC/0.1% SDS at room temperature for 15 minutes, and twice in 0.1X SSC/0.1% SDS at 65-69^{0}C for 30 minutes. The blots were dried on paper towels, wrapped in Saran wrap, and exposed for 1 to 7 days to Kodak XAR-5 film backed by an intensifying screen at -70^{0}C.

Restriction fragments of the cosmid were also labeled and used as probes on Southern blots. They were cut out from

low-melting-point (LMT) agarose gels, mixed with water at 2 ml per gram of gel, heated at 100^{0}C for 10 minutes, and oligolabeled as previously described without further purification. Plasmid subclones were also oligolabeled and used as probes on Southern blots.

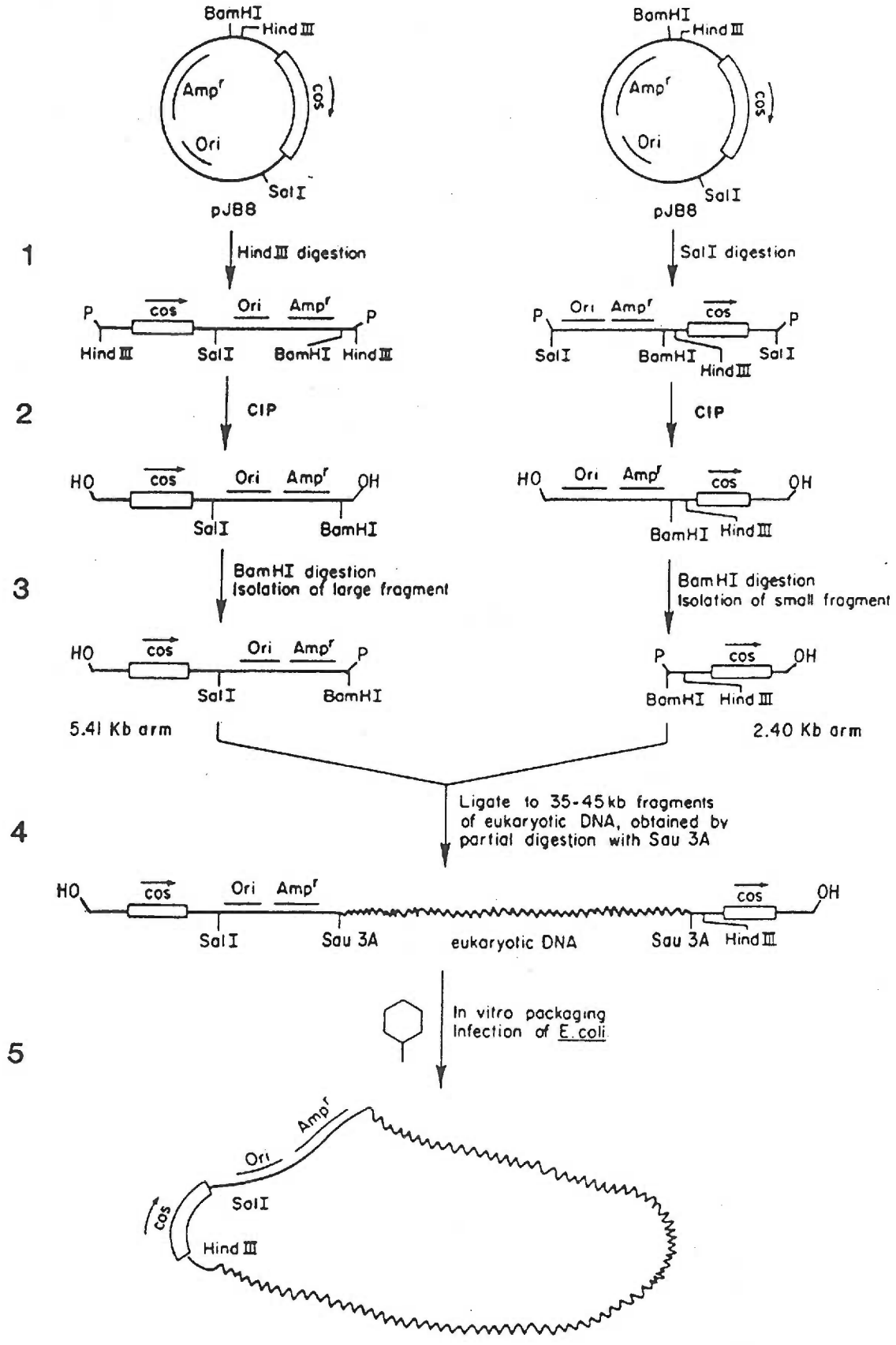
4. Somatic cell hybrid mapping panel. The mapping panel consisted of 23 cell lines provided by Gail Bruns, Children's Hospital, Boston. The G35 cell lines are human-hamster hybrids derived from fusion of the Chinese hamster cell line E-36 with white blood cells from a female carrier of the X/19W translocation t(X;19)(q23-25::q13)(Latt et al, 1976). The G17 and G24 cell lines are human-mouse hybrids derived from fusion of the mouse cell line RAG with white blood cells of the X/19W translocation carrier (G17 lines) or the X/19B translocation t(X;19)(q1::p13)(Brook et al, 1978). The G89E5 and G95A4 cell lines were derived from fusions of white blood cells from a karyotypically normal male with the Chinese hamster cell lines YH21 and Wg3h, respectively. G35A5AzA and G35A5AzF are subclones of G35A5 selected in 8-azaguanine for loss of the der 19 translocation chromosome. These hybrid cell lines were characterized by both isozyme and cytogenetic techniques (Bruns et al, 1978; Bruns et al, 1979). In addition, DNAs from these hybrids have also been analysed with cloned DNA probes for all chromosomes except the Y.

5. In situ hybridization. Probes p79-2-23 and CF 33-79 were nick-translated according to the method of Harper and Saunders (1981) to a specific activity of 4 x 10^7 dpm/µg using [^{3}H] TTP (65

Ci/mmol) and [^{3}H]dCTP (60 Ci/mmol)(Amersham). p79-2-23 and CF33-79 were then prehybridized with total human DNA as previously described. In situ hybridization to metaphase spreads from normal male cells was performed according to Harper and Saunders (1981) by B. Sheehy in E. Magenis' laboratory. This method is described in the section III manuscript.

6. Terminology for cosmids and plasmids. Cosmids were named beginning with CF, to designate the somatic cell hybrid line, followed by the plate number and the clone number on that plate. For example, a clone from the second well in row C of 96-well plate 9 was called CF9C2. A cosmid from master gridded plate 33 at square 79 was called CF33-79.

The plasmids were designated by a small p followed by the cosmid square number, the subclone master plate number, and the square number of the subclone plate. For example p79-2-23 was a subclone of cosmid CF33-79 from the 23rd square of subclone master plate 2.


BamHI
Hind III
Amp^r
cos
Ori
Sal I
pJB8
1
Hind III digestion
Sal I digestion
P
Ori
Amp^r
Hind III
Sal I
BamHI
2
CIP
HO
OH
3
BamHI digestion
Isolation of large fragment
BamHI digestion
Isolation of small fragment
5.41 Kb arm
2.40 Kb arm
4
Ligate to 35-45 kb fragments
of eukaryotic DNA, obtained by
partial digestion with Sau 3A
Sau 3A
eukaryotic DNA
5
In vitro packaging
Infection of E. coli

B. Results

1. High molecular weight DNA. Figure 2 shows a 0.3% agarose gel containing uncut bacteriophage λ DNA and CF52-46-1/8 DNA from a high molecular weight preparation. The size of such large DNA cannot be accurately measured, but as very high molecular weight DNA does not separate well, DNA migrating well above the 50 kb λ DNA, as in this case, is likely greater than 200 kb in size (Maniatis et al, 1982).

After digestion and sucrose gradient size fractionation, 15 fractions were collected, and aliquots from every third fraction were checked for size on a 0.4% agarose gel. Fractions 5 through 8 contained DNA in the 35 to 45 kb size range and were pooled for use as insert DNA in subsequent ligations.

2. Ligation, *in vitro* packaging, and transduction into bacterial hosts. Figure 3 shows 2 ligation test gels. Ligation is indicated by an increase in the size of the insert DNA smear due to concatamerization, by the partial disappearance of the 5.4 kb and 2.4 kb vector arms, and by the appearance of vector-vector ligation products of 10.8 kb (2 left arms), 7.8 kb (left arm + right arm), and 4.8 kb (2 right arms).

The efficiency of ligation was not measured. It could not be determined by gel monitoring of the appearance of high molecular weight DNA as the insert DNA was already very large. In fact, King & Blakesley (1986) found that efficiency does not correlate with the

appearance of a very high molecular weight DNA. It could not be determined using transduction of *E. coli* as the DNA was first packaged *in vitro* adding another variable that influenced the efficiency of transduction. While the insert DNA shown in figure 3b did not increase as noticeably in size as the DNA in figure 3a, both gave similar numbers of transformants after packaging and transduction.. As a complete library was not being constructed, maximum efficiency of ligation and packaging was not critical.

Initial transductions of the packaged ligation mix were done with *E. coli* strain ED8767. Efficiencies ranged from 3000 to 8000 colonies per microgram (µg) of total DNA. A total of about 2000 colonies from these transductions were transferred to either 96-well plates in freezing medium or to a gridded agar plate for storage and preparation of replica filters.

Later transductions were done using HB 101 with efficiencies of 3300 to 9000 colonies per µg of total DNA. These colonies were larger and appeared to grow better than those of ED8767. A total of about 1500 colonies were transferred to gridded agar plates for storage and preparation of replica filters.

One transduction was done with *E. coli* strain DK-1. The colonies were extremely small, efficiency was poor (800 colonies/µg total DNA), and the viability was poorer than for the other strains.

3. Screening the library for clones with human inserts. The cosmid screening results are shown in table 1. From 3400 clones, one would expect at least 85 clones, or 2.5%, to contain human

inserts based on the chromosome composition of the somatic cell hybrid line used to make the library. On screening of replica filters with total mouse and total human DNA, however, only 50 of 3400 clones, or 1.5%, appeared positive for human DNA and negative for mouse DNA. Figure 4 shows two autoradiographs from the screening of plate 33. Filter 33a was probed with total human DNA; 33b with total mouse DNA. (Screenings were actually done in duplicate, i.e., two filters were probed with total human and two with total mouse.) Note that only 6 clones that light up with the total human probe, 79, 157, 176, 242, 372, and 393, light up faintly or not at all with the total mouse probe. A number of clones hybridized with both mouse and human probes.

4. Preparation of cosmid DNA. The next step was to prepare cosmid DNA from those clones that appeared to contain human inserts. Using the Birnboim alkaline miniprep method (Birnboim, 1983), yields were consistently poor compared to those obtained with known plasmids and cosmids. Numerous modifications were tried with little improvement. Figure 5a shows the results of a typical series of preparations from 4 ml of overnight culture of clones stored at -70^{0}C in freezing medium as checked on a 0.45% agarose gel. A known control cosmid (not shown), 1-13, gave yields of about 1-2 µg per ml of fresh overnight culture while DNA from the test cosmids gave yields of 0-30 ng/ml. Lane 1 is 187 ng of cosmid 1-13 from a previous preparation to serve as a size standard and as a quantitative comparison. Preparations made from fresh

overnight cultures from the -20^{0}C gridded plates gave somewhat better, but still poor and highly variable, yields as shown in figure 5b. Yields ranged from unmeasurable to about 1 µg/ml of culture. I was unable to get any measurable amount of DNA from 7 of the 50 cosmids.

During these initial minipreps, 5 of the 50 cosmids were found to be much smaller than 35-45 kb, having apparently deleted most of their DNA to less than 20 kb. They must have been larger initially as cosmids less than about 35 kb would not have packaged into bacteriophage λ (Old & Primrose, 1985). Figure 5b shows one such deleted cosmid, CF32-131, in lane 2; figure 6 shows a second one, CF25-350, in lane 2.

In addition to the alkaline minipreps, large scale preparations including purification in cesium chloride (CsCl) gradients were attempted. Once again poor or unmeasurable yields resulted. In fact, in most cases no cosmid DNA band was visible in the gradient after overnight ultracentrifugation. Only one intact cosmid, CF33-79, gave a measurable yield of17.5 µg from 350 ml of culture. I got a surprisingly good yield of 90 µg from 175 ml of culture containing cosmid CF33-176 only to discover that the cosmid had deleted to several forms (figure 6), all less than 4 kb, in size indicating that they had lost part of the vector as well as the human insert. Another cosmid, CF 32-391, with a yield of 8µg of cosmid DNA from a 175 ml culture had deleted to about 6 kb.

Intact cosmid DNAs isolated from several HB 101 host clones were repackaged and transduced into DK-1, a more stringent Rec A^- host strain due to a deletion of the Rec A gene (D. Kurnit, pers. comm., 1985), in an attempt to improve DNA yields and eliminate the deletion problem (Gumucio et al, 1985). Neither goal was accomplished. CsCl gradient preparations were still unsuccessful, and the cosmids lost DNA even more rapidly than those in HB 101. For example, CF32-391 DNA minipreps from fresh 5 ml overnight culture inoculated from a -70^0C HB 101 clone yielded consistently intact cosmid DNA while a similar preparation of this cosmid made from a -70^0C DK-1 clone yielded only severely deleted cosmid DNA. Minipreps of cosmid CF33-176 DNA gave similar results.

5. Screening cosmids for polymorphisms. A total of 42 cosmids were labeled by nick translation, prehybridized with total human DNA, and screened on Southern blots containing Taq I and Msp I digested human DNAs from 6 to 9 unrelated individuals. Overall, 23 of the cosmids did not hybridize to any human DNA, i.e. the blots were blank except for the size standards. Two cosmids showed heavy lane background, 9 cosmids revealed nonpolymorphic fragments, and 8 cosmids revealed variable fragments that possibly represented polymorphisms. Six of the cosmids that did not hybridize to human DNA were later hybridized on Southern blots that included a lane of DNA from the somatic cell hybrid line used to make the library. In all cases, these probes hybridized strongly to the somatic cell hybrid DNA lane and not at all to the random human

DNAs, indicating that these cosmids contained mouse inserts rather than human inserts. These 6 clones had initially been chosen as containing human inserts even though they had hybridized only very faintly with the total human DNA on colony screening because they did not appear to hybridize at all with the total mouse probe.

6. Further screening of cosmids revealing polymorphisms. Of the 8 cosmids that showed variable size fragments when hybridized to Taq I and Msp I digested total human DNAs, four showed enough variability to make additional screening worthwhile: CF32-391, CF32-435, CF33-79, and CF33-176. Autoradiographs from the initial screening of CF32-391, CF33-79, and CF33-176 are shown in figures 7, 8, and 9. CF32-435 showed a fragment pattern identical to that of CF 33-79 indicating that these 2 cosmids contained the same or an overlapping human insert.

CF32-391 revealed variable Taq I fragments of >18 kb and 6.6 kb (figure 7). Cf33-79 revealed variable Taq I fragments of 6.4, 3.7, 3.5, 3.2, and 3.0 kb (figure 8), and variable Msp I fragments of 1.5 and 1.1 kb (Msp I blot not shown). CF33-176 revealed variable Taq I fragments at 7 kb and variable Msp fragments at 5.5 and 3 kb (figure 9).

7. Cosmid CF33-79. The next step was to find single copy subclones of the cosmid that revealed specific polymorphisms. CF33-79 was studied first as it looked most promising, and it was the only one of the 3 cosmids for which a reasonable amount of DNA

(17.5 µg) was recovered. The approach was to cut the cosmid with various 6-bp recognition sequence enzymes to find one that gave several well separated, moderate sized fragments so that these fragments could be recovered from LMT agarose and used as probes on Southern blots of Taq I, Msp I, and Rsa I digested total human DNAs in hopes of finding a single fragment that showed a polymorphism. At the same time, the cosmid was digested with Sau 3A and the fragments cloned into pSP65 to obtain about 200 subclones of the cosmid. The fragments revealing polymorphisms were used as probes of colony filters containing the subclones. Duplicate filters were also probed with total human DNA to eliminate subclones containing repetitive DNA. Subclones that hybridized to a fragment of interest but not to the total human DNA were chosen for further study by growing up the clones, isolating the plasmid DNAs, and using them as probes on the same or similar Southern blots as used above.

CF33-79 was digested with Eco RI, Kpn I, Sac I, and Hind III. Eco RI gave the best size range with five well separated fragments of 25-30, 7.5, 5.8, 2.7, and 2.25 kb. When these fragments were oligolabeled, prehybridized with total human DNA, and used as probes on Southern blots, the 25 and 2.7 kb fragments hybridized to nonpolymorphic bands, the 2.25 kb fragment showed only faint lane background, and the 7.5 kb fragment revealed the Rsa I and Taq I polymorphisms shown in figure 10. The 5.8 kb fragment revealed the same Taq I polymorphism but hybridized only very weakly with the

human DNA and very strongly with the size standards. This finding indicates that the 5.8 kb fragment includes the junction of the human insert and the vector.

Sau 3A fragments of CF33-79 were subcloned into the Bam HI site of pSP65, competent HB 101 were transformed with the ligation mix, and colonies containing recombinant plasmids were selected on ampicillin plates. From these plates 2 master gridded plates containing about 100 colonies each were made. Replica filters from these plates were probed with oligolabeled total human DNA and with the oligolabeled 7.5 kb Eco RI cosmid fragment. Both hybridizations were done in duplicate. Thirteen subclones hybridized to the cosmid fragment but not to total human DNA. DNA preparations were made from these clones and from 6 others with inconclusive hybridization results. Of these 19 subclones, 5 had very small or no inserts and were not studied further. The remaining 14 subclones were hybridized with Southern blots as described above with the following results: 6 contained repeats so that single bands were not visible, 2 revealed nonpolymorphic bands, 3 produced blank or uninterpretable autoradiographs, and 3 (p79-1-98, p79-1-103, & p79-2-23) showed the same polymorphisms revealed with the 7.5 kb cosmid fragment.

The Taq I and Rsa I polymorphisms appear to represent the same locus, that is, an insertion/deletion RFLP. The same locus was also observed when p79-1-103 was used on Southern blots of Hind III, Pvu II, Msp I, and Eco RI digested total human DNAs. This correlation between genotypes with 6 different enzymes is strong

evidence for an insertion/deletion model. The equivalent size difference between allelic fragments found with Rsa I and Taq I, as shown in table 2, is further evidence. The higher molecular weight allele sizes could only be estimated for Rsa I as 4.36 kb was the largest size standard on the blots whereas they were determined more accurately on Taq I blots which contained an 8.5 kb size standard.

Figure 11, an autoradiograph of an Rsa I blot of eighteen individuals probed with oligolabeled p79-2-23, shows the extensive variation at this locus among unrelated individuals. A total of 51 unrelated individuals were genotyped at this locus; 41 (80%) were heterozygous.

The RFLP was then studied in six 3-generation families, and Mendelian inheritance was observed in all cases. These six pedigrees are shown in Appendix A. Table 3 summarizes the parental genotypes in these plus 3 additional families. Figure 12 shows the Mendelian inheritance of the RFLP in two families, K1331 and K1340.

a. Chromosomal localization of cosmid CF33-79. Initially, the whole nick-translated cosmid, prehybridized with total human DNA, was used to probe Southern blots of the somatic cell hybrid mapping panel. The blots were poor and inconclusive. The blots were also probed with CF32-435, and while the quality was poor, there was an indication that the cosmid was on chromosome 16 as it hybridized to cell lines G35F3, G35E3, G35C2, G35F5, G35C5, G35A4, G24B5, and RRP3-6B1. The blots were later probed with the human insert from

subclone p79-2-23; the results are shown in table 4. There was 100% concordance only with chromosome 16. (The autoradiograph is not shown, but the probe hybridized to the same cell lines as CF32-391 subclone p391-2-42 shown in figure 16.) There were 8 discordancies with chromosome 11; four were cases in which the probe hybridized to cell lines that do not contain chromosome 11.

A second chromosome localization method, in situ hybridization to metaphase chromosomes, was done by B. Sheehy in E. Magaenis' laboratory using [^{3}H] labeled p79-2-23 and cosmid CF33-79, both of which were first prehybridized with total human DNA. This method also localized the RFLP to chromosome 16, specifically to the distal half of 16q as shown in Figure 13. Thirty of 150 cells scored (20%) had a grain localized to the region 16q13->16qter. There appeared to be a second peak on chromosome 7, with 5% of cells showing a grain localized to the proximal part of 7p. This may represent a second locus, however, such an idea is not supported by the somatic cell hybrid mapping panel results of 6 discordancies between chromosome 7 and p79-2-23.

8. Cosmid CF32-391. Because large scale DNA preparations using CsCl gradients were unsuccessful, nineteen 1.65 ml minipreps were done and combined for a total yield of about 6.5 µg (200 ng/ml culture) of what appeared to be intact cosmid DNA. The approach used with CF33-79 to find single copy fragments revealing polymorphisms was initially used with CF32-391 except that the

DNA was cut with only one 6-bp recognition sequence enzyme, EcoRI, due to the small amount of DNA available. The EcoRI digests gave 13 fragments ranging in size from 5.5 kb to less than 1 kb. Some fragments were too closely spaced to isolate separately from LMT agarose and were cut out of the gel together. The 6 largest fragments were oligolabeled, prehybridized with total human DNA, and used as probes on Southern blots of Taq I, Msp I, and Rsa I digested human DNAs. The 5.5 kb fragment revealed an extra 6.5 kb Msp I fragment in one of 8 individuals. A 5 kb cosmid fragment revealed an extra 7.5 kb Taq I fragment in the same individual, and a 3 kb fragment revealed an extra 3.6 kb Msp fragment in the same individual. The high molecular weight (MW) Taq I polymorphism seen when the whole cosmid was used as a probe was not seen with any of the cosmid fragments. One explanation is that high molecular weight fragments do not transfer well on Southern blotting and were absent or poorly represented on the blots probed with the cosmid fragments.

Sau 3A fragments of CF32-391 were subcloned into pSP65 as previously described and replica colony filters screened with total human DNA and with the 3 cosmid fragments revealing possible polymorphisms. Of 166 clones, 43 hybridized with total human DNA. Unfortunately, when the colony filters were probed separately with the 5.5, 5 and 3 kb cosmid EcoRI fragments, all of the fragments hybridized to all of the clones. It is likely that these fragments contained vector DNA, either as part of the fragment or from a

nearby contaminating fragment (as gel separation was poor) which hybridized with homologous regions of the subclone vector.

The next approach was to look for the Taq I and Msp I polymorphisms through mass screening of the subclones that did not hybridize to total human DNA. DNA was prepared from 95 subclones; 43 had reasonably sized inserts (>400 bp) and were oligolabeled and screened in groups of 1 to 6 on Southern blots of Taq I and Msp I digested human DNAs. None of the screened plasmids revealed the Msp I or small Taq I polymorphisms. Finally, in the last group to be screened, p391-2-42 revealed the high MW Taq I polymorphism.

Figure 14a is an autoradiograph of the probe hybridized with Taq I digested DNA from 18 individuals. The probe hybridized poorly to these large fragments compared to its hybridization to the constant 3.9 kb fragment. This finding is probably partly due to poor transfer of high MW DNA onto the Southern blots; it may also indicate that the probe is highly homologous with the 3.9 kb fragment and only partially homologous to the larger fragments.

It was difficult to get good size estimates of such large, closely spaced fragments as most were larger than the largest size standard fragment of 18.4 kb. One individual (695) had three variable fragments indicating that these fragments may represent more than one locus. The probe was hybridized to DNAs on a Taq I titration blot and on 6 family blots to rule out partial digestion and to determine the inheritance pattern of the fragments.

The high molecular weight fragments were present on the titration blot at all 3 enzyme concentrations, 2X (2u enzymes/µg DNA), 10X, and 20X (not shown). The fragments appeared to conform to Mendelian inheritance in four families, one of which, K1340, is shown in figure 14b. In two families, however, the fragment patterns were too complex to be explained by a single locus. Some children in each family had 3 fragments, and although each fragment was present in one or both parents, the inheritance pattern was not clear. These studies were hampered by the poor transfer and hybridization of these large fragments, and even with film exposure times exceeding 7 days, the hybridizations were too weak in some individuals to determine the fragment patterns with certainty.

a. Chromosomal localization of cosmid CF32-391. Plasmid p391-2-42 was used as a probe on Southern blots of the somatic cell hybrid mapping panel; the Autoradiograph is shown in Figure 15. As with p79-2-23, there was 100% concordance only with chromosome 16 (Table 4). There were 8 discordancies with chromosome 11. A haptoglobin (a gene known to be on 16q) probe hybridized to the same cell lines when used on the mapping panel.

In regard to the great variability in the hybridization signal among the positive cell lines, through extensive use of these cell lines in this laboratory, similar variability has been observed with many other probes.

9. Cosmid CF33-176. As with cosmid CF32-391, a series of minipreps were done and gave a total yield of about 5 µg of intact cosmid DNA from 17 ml of fresh overnight culture (300 ng/ml). Digestion with Eco RI produced 12 fragments, some poorly separated, ranging in size from 10 kb to less than 1 kb. The largest 7 fragments were cut out of LMT agarose, oligolabeled, prehybridized with total human DNA, and used to probe Southern blots of Taq I and Msp I digested human DNAs. The 10 kb Eco RI cosmid fragment revealed a variable 6 kb Msp I fragment in one individual, and the 5 kb Eco RI cosmid fragment revealed a variable 7 kb Taq I fragment in two individuals.

Sau 3A fragments of the cosmid were subcloned into pSP65 and colony filters containing 178 recombinant plasmids were probed with total human DNA and with the 10 kb and 5 kb cosmid Eco RI fragments in search of single copy probes that would reveal the polymorphisms found with the whole cosmid and with the Eco RI fragments of the cosmid.

Nineteen of the clones hybridized strongly with total human DNA. As with cosmid CF32-391, the CF33-176 EcoRI fragments hybridized with just about all of the subclones. However, the fragments showed great variability in the degree of hybridization to different subclones so that 24 showing the greatest intensity with the 10 kb EcoRI fragment and 21 showing the greatest intensity with the 5 kb fragment (but no hybridization with total human DNA) were selected for further study. After plasmid DNA preparation and

elimination of plasmids containing tiny or no inserts, 20 plasmids were used as probes on Msp I digested human DNAs and 18 on blots of Taq I digested total human DNAs.

Subclone p176-1-65 revealed what appeared to be a 2-allele RFLP with variable Msp fragments of 4.65 and 5.85 kb. Subclone p176-1-108 revealed what appeared to be two 2-allele loci, one with variable Msp I fragments of 5.6 and 5.45 kb and a second with variable Msp I fragments of 2.2 and 1.8 kb. Subclone p176-2-20 revealed a possible 2 allele locus with variable Msp I fragments of 5.6 and 5.2 kb. Interestingly, the first, third and fourth of these loci appeared to be in complete linkage disequilibrium, i.e. an individual heterozygous at one locus was heterozygous at the other 2 loci. Also, no homozygotes for the larger of the 2 alleles was seen; all individuals were either heterozygous or homozygous for the smaller of the two fragments. This was also true of the other locus (5.6 and 5.45 kb) which was not in complete disequilibrium with the other three. Finally, heterozygosity was low; 22% for the 5.85/4.65 kb locus and 17% for the 5.6/5.45 kb locus.

As Msp I is notorious for giving incomplete digests at some sites in human DNA even at high enzyme excess (Busslinger et al, 1983), it was important to verify that these possible polymorphisms were not a result of incomplete digestion. For this purpose, a titration blot was probed with oligolabeled p176-1-108. This blot contained DNAs from 3 individuals, each digested with 3 increasing concentrations of Msp I of 1X (1 unit/µg DNA), 5X, and 10X. The

extra fragments at 5.6 and 2.2 kb were still visible at 10X but hybridize with less intensity indicating that they may be the result of partial digests.

The second check was to probe an Msp I blot of family K1345 to see if the fragments were inherited in a Mendelian fashion. The autoradiograph of this hybridization is shown in figure 16. One can see that the fragments are definitely not inherited a la Mendel. One child shows the 2.2 kb fragment while none of the parents or grandparents have it; several children and 2 grandparents have the 5.6 kb fragment although neither parent has it.

From these studies, it appears that these "RFLPs" are in fact artifacts due to partial Msp digestion of human DNAs. They probably represent Msp I sites that are particularly resistant to cleavage, as has been found for GGC^{m}CGG (Busslinger et al, 1983).

None of the subclones used on Taq I digested human DNAs revealed the Taq I polymorphism seen with the whole cosmid and with the 5 kb Eco RI cosmid fragment.

TABLE 1

Cosmid Screening Results

Total colonies screened with total human and total mouse DNA	3400
Clones initially positive for human inserts	50
Cosmids with successful DNA preparations	43
Cosmids with obvious deletions	5
Cosmids oligolabeled and screened on Taq I and Msp I digested human DNAs	43
Cosmids with no detectable hybridization to human DNAs	23
Cosmids with repeats that obscured visualization of individual fragments	2
Cosmids revealing nonpolymorphic fragments	10
Cosmids revealing possible polymorphisms	8

TABLE 2

CF33-79 RFLP*

Fragment pattern (kb)		Allele frequency
Rsa I	Taq I	(102 chromosomes)
7.1**	7.3	0.01
6.2**	6.4	0.08
5.9**	6.1	0.13
5.1**	5.3	0.02
4.8**	5.0	0.03
4.5	4.7	0.01
3.9	4.1	0.05
3.5	3.7	0.19
3.3	3.5	0.04
3.0	3.2	0.37
2.8	3.0	0.08

*This RFLP is seen with three subclones of CF33-79: p79-1-98, p79-1-103, and p79-2-23 with insert sizes of 1.45 kb, 3.0 kb, and 1.45 kb, respectively.

** Estimated size based on size of Taq I fragments.

TABLE 3

CF33-79 RFLP

Parental Genotypes for Taq I fragments (kb)

Family	Father	Mother
K1329A	3.2, 3.0	3.2, 3.2
K1329B	3.7, 3.2	6.1, 3.2
K1329C	3.2, 3.2	3.7, 3.0
K1329D	6.1, 3.2	3.2, 3.2
K1331	6.1, 3.7	3.7, 3.5
K1333	6.1, 6.1	6.4, 3.2
K1340	3.7, 3.2	5.0, 3.2
K1341	6.1, 3.2	5.0, 4.1
K1345	3.2, 3.2	6.4, 3.2

TABLE 4

Chromosome contents of hybrid cell lines. The designations are: (+) presence or (-) absence of a human chromosome; (R), rearranged chromosome as determined by disruption of a syntenic group or by a cytogenetic abnormality; (±), chromosome present in less than 15% of cells and/or the isozyme or DNA probe characteristic of the chromosome weakly positive; (p), presence of a marker for the short arm and absence of a marker for the long arm; (q), presence of a marker for the long arm and absence of a marker for the short arm; (a), Xq24→qter. For calculation of the discordancy fractions, hybrids with a rearranged chromosome or those where the chromosome was present in less than 15% of the cells were excluded from analysis.

The column designated M indicates the presence or absence of human-specific Hind III fragments on Southern blots probed with p79-2-23i and Taq I fragments on Southern blots probed with p391-2-42i.

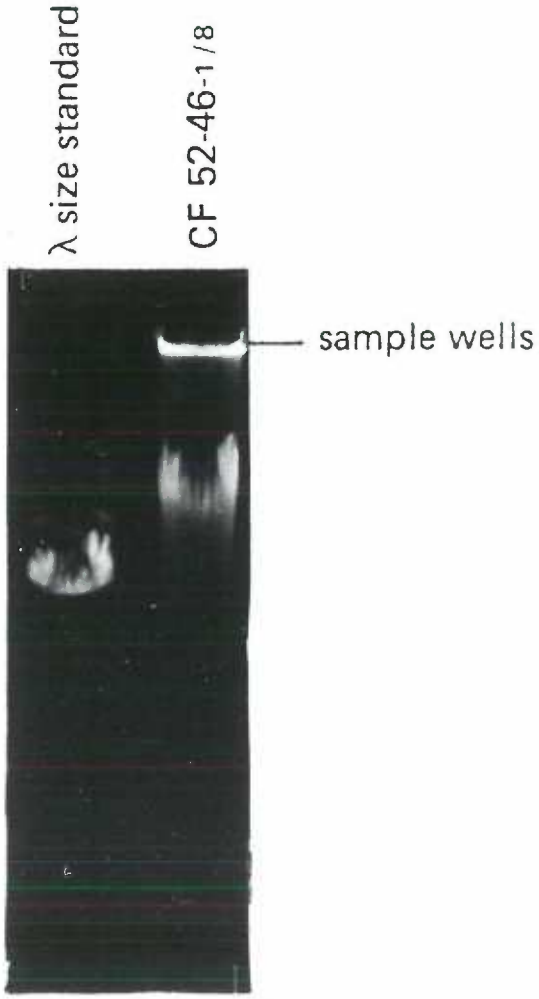
[b]Cell line G89E5 contains an X chromosome in all cells plus an unidentified human-hamster translocation in about 20% of the cells.

Cell Line	Chromosome Complement 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	X	Y	Scoring M
G35D5	+	+	+	−	−	+	+	−	−	+	−	−	+	+	+	+	+	+	+	+	−	+	a	−	+
G35F3	−	+	+	−	+	−	+	+	−	+	+	+	−	−	−	+	−	+	+	+	−	−	a	−	+
G35D2	±	−	−	−	+	+	−	−	+	−	+	+	−	−	+	−	−	+	R	+	−	+	aR	−	−
G35E3	−	−	−	−	+	+	−	+	−	−	+	+	+	+	−	+	−	+	+	+	+	+	a	−	+
G35A2	−	−	+	±	−	+	−	−	−	−	+	−	−	+	−	−	−	−	+	+	+	+	+	−	−
G35C2	−	−	−	+	−	−	−	+	−	−	R	−	−	+	−	+	−	−	R	−	−	−	R	−	+
G35C4	−	−	−	−	−	+	−	−	R	−	−	+	−	+	−	−	−	±	+	+	+	−	a	−	−
G35B5	−	−	−	−	−	R	+	−	−	−	+	p	+	−	−	−	+	−	+	+	−	−	a	−	−
G35F5	±	−	+	+	−	+	−	−	+	−	+	−	+	+	−	+	+	+	+	±	+	+	+	−	+
G35D3	−	p	−	±	−	−	+	+	+	+	−	−	−	±	−	+	+	−	+	+	+	+	a	−	+
G35C5	−	−	p	−	+	−	+	+	+	+	−	p	−	−	+	+	+	−	+	+	+	+	a	−	+
G35A4	+	−	+	+	−	+	−	+	+	−	−	−	−	+	−	+	−	−	+	+	+	−	a	−	+
G35B4	+	−	−	−	+	−	−	−	−	+	+	+	−	−	−	−	−	+	+	−	−	+	+	−	−
G35E4	−	−	−	R	−	−	−	−	R	−	−	−	−	−	+	−	−	−	+	−	−	−	a	−	−
G175	±	−	−	+	−	+	−	+	−	+	−	−	+	+	+	−	−	−	+	−	−	+	a	−	−
G24A9	−	+	−	+	−	+	−	+	+	+	−	−	±	+	+	−	±	+	+	−	+	−	q	−	−
G24A4	±	p	−	−	+	+	+	−	±	−	±	−	+	+	+	+	−	−	+	−	+	−	q	−	+
G24B5	+	−	+	+	−	−	+	−	−	+	+	+	+	+	−	+	−	−	+	−	+	+	+	−	+
G89E5[b]	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	−	−
G95A4	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	±	−	−	+	−	−
RRP5-7	+	−	−	+	+	−	+	−	−	+	−	+	ND	+	−	−	+	+	−	+	−	−	+	R	−
RRP3-681	+	+	+	+	+	+	+	−	−	+	+	+	−	−	+	+	+	+	R	−	+	−	+	+	+
G35F1	−	−	−	+	+	−	−	+	−	±	−	p	−	+	−	−	−	+	+	+	+	−	a	−	−
Discordancy Fraction	7/19	7/21	5/22	9/20	10/23	9/22	6/23	8/23	8/20	9/22	8/21	10/20	8/21	9/22	11/23	0/23	8/22	11/22	8/20	10/21	7/23	9/23	5/8	11/22	

FIGURE 2

High molecular weight CF52-46-1/8 DNA preparation. The λ DNA is about 50 kb.

High Molecular Weight CF52-46-1/8 DNA Preparation



0.3% agarose gel

FIGURE 3

Ligations. Ligation is indicated by an increase in the size of the insert DNA smear due to concatamerization, by the partial disappearance of the 5.4kb and 2.4kb vector arms, and by the appearance of vector-vector ligation products.

Ligations

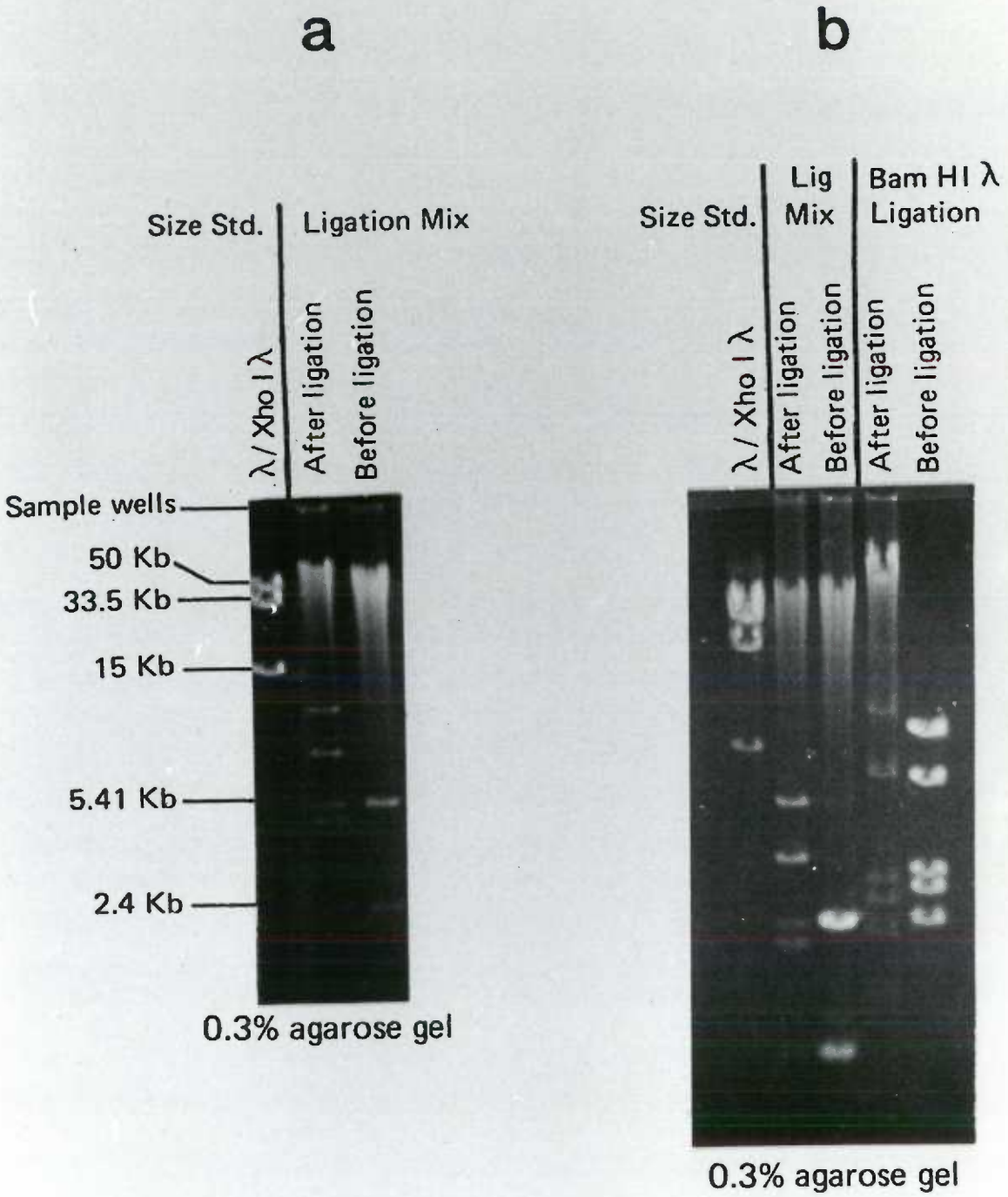
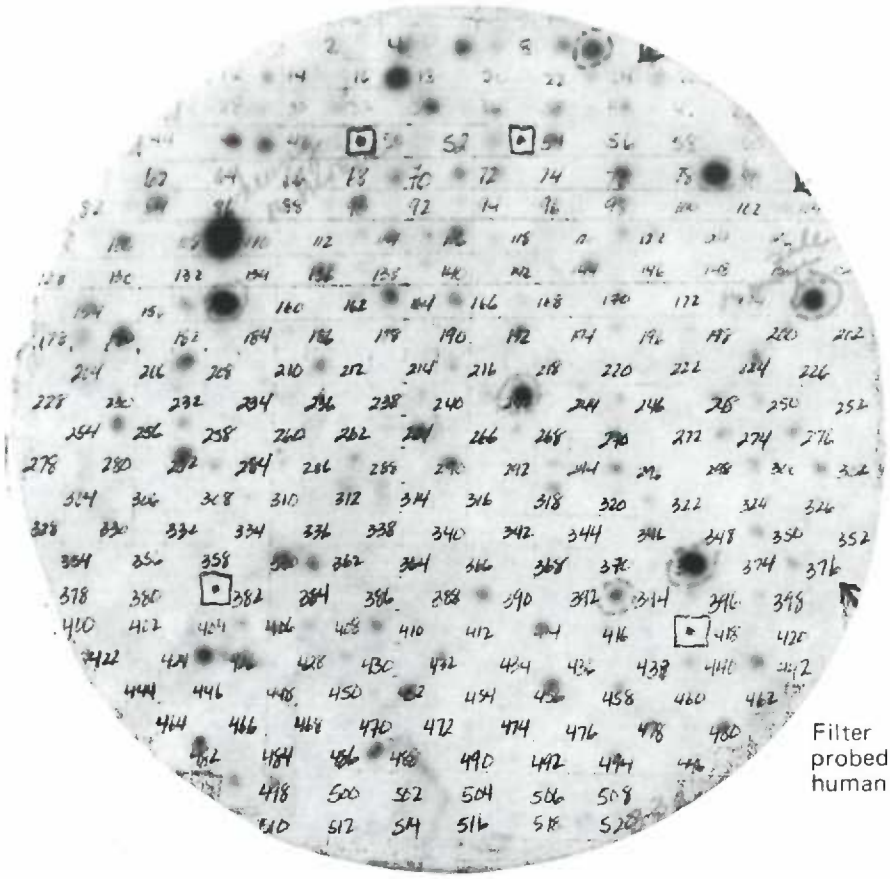
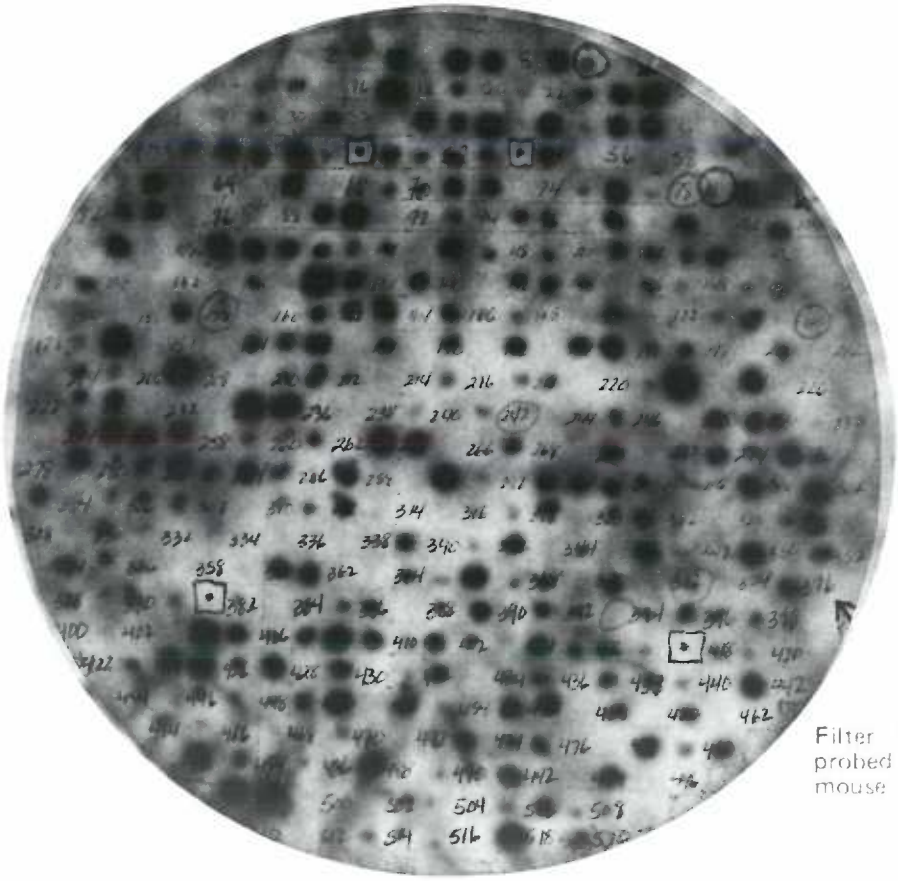

a
b
Size Std.
Ligation Mix
λ / Xho I λ
After ligation
Before ligation
Sample wells
50 Kb
33.5 Kb
15 Kb
5.41 Kb
2.4 Kb
0.3% agarose gel
Size Std.
Lig Mix
Bam HI λ Ligation
λ / Xho I λ
After ligation
Before ligation
After ligation
Before ligation
0.3% agarose gel

FIGURE 4

Colony screening of cosmid library plate 33. Six clone hybridized strongly with total human DNA and weakly or not at all with total mouse DNA: 79, 157, 176, 242, 372, and 393.



Filter 33a probed with total human DNA



Filter 33b probed with total mouse DNA

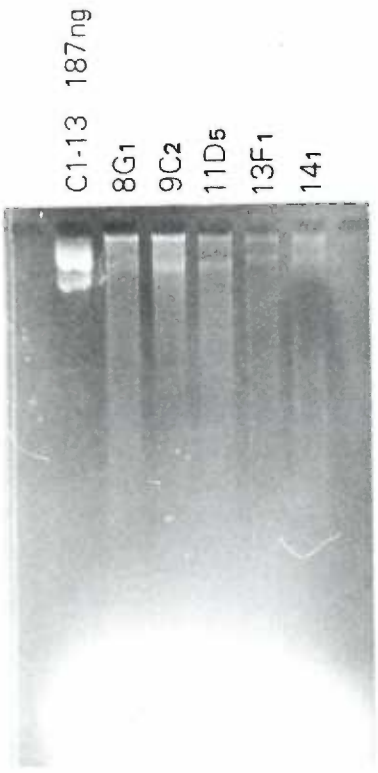
FIGURE 5

Cosmid DNA preparations. The Birnboim alkaline miniprep method was used (Birboim, 1983). Cosmid 1-13 served as a size standard and quantitative comparison. Cosmid CF32-131 has deleted most of its DNA.

Cosmid DNA Preparations

a

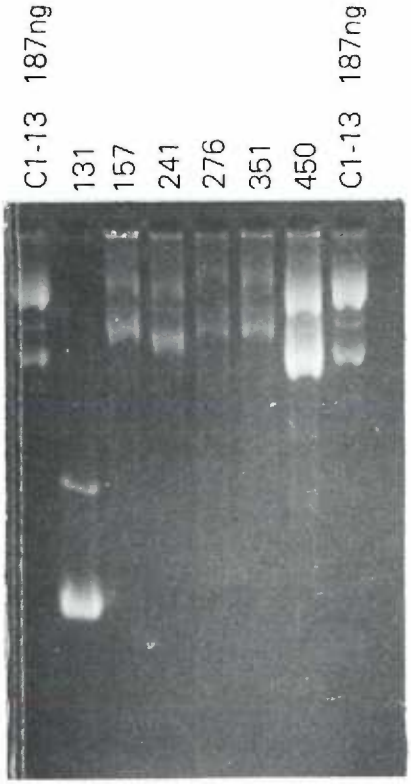
Cosmids from freezing medium



0.45% agarose gel

b

Plate 32 cosmids



0.4% agarose gel

FIGURE 6

Deleted cosmids. CF25-350 showed a single deleted form on minipreparation and multiple deleted forms, as shown here, after large scaled preparation.

CF33-176 was intact cosmid DNA from minipreps of HB101 clones. It was obviously deleted after minipreps and large scale preparation, as shown here, when cloned into DK-1.

Cosmid 1-13 and Hind III-digested λ serve as size standards.

Deleted Cosmids

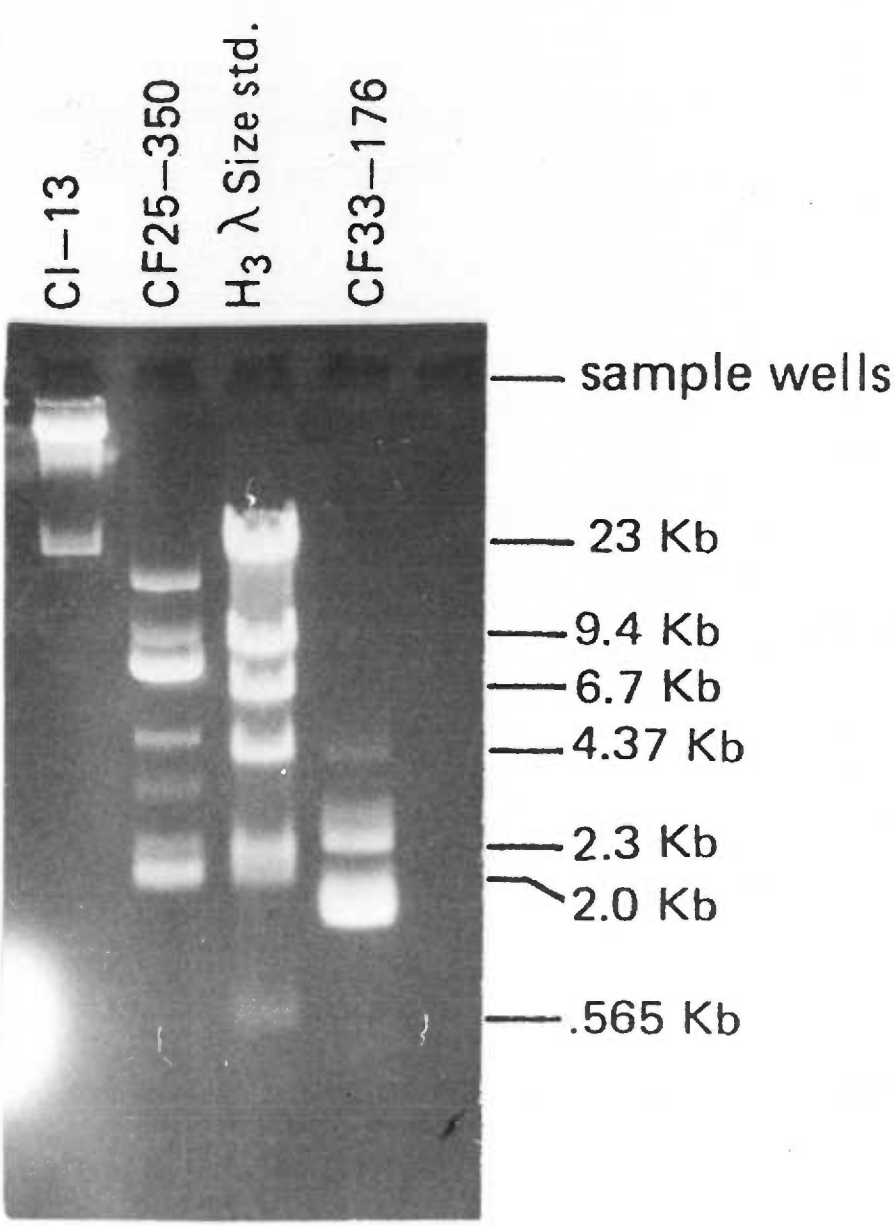


FIGURE 7

Southern blot of Taq I digested DNA from 6 unrelated individuals probed with cosmid CF32-391. The probe was first prehybridized with total human DNA to prevent repetitive DNA in the cosmid from hybridizing to homologous repetitive DNA on the Southern blots.

The probe hybridized to variable fragments above the 18.4 kb size standard in all individuals and to a 6.6 kb fragment in one individual (lane c). It also hybridized to several smaller constant fragments of 3.9, 3.5, and 3.3 kb (shown) and 2.4 and 1.9 kb (not shown). Lines to the left of the figure indicate variable fragments.

Lane g contains size standards.

CF32-391

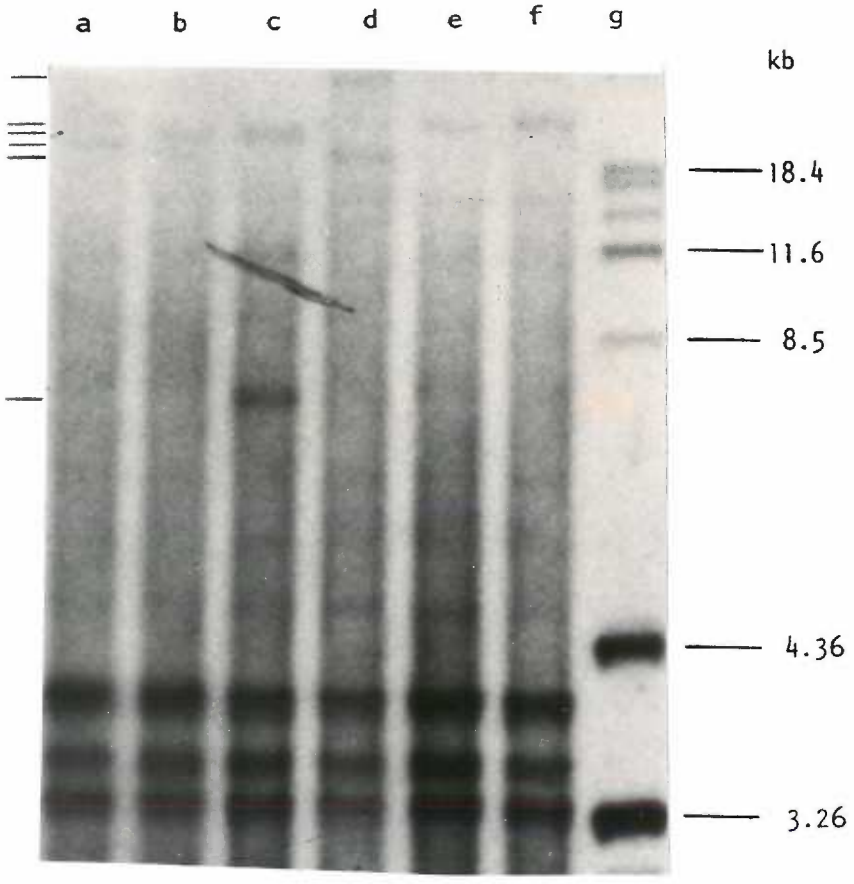
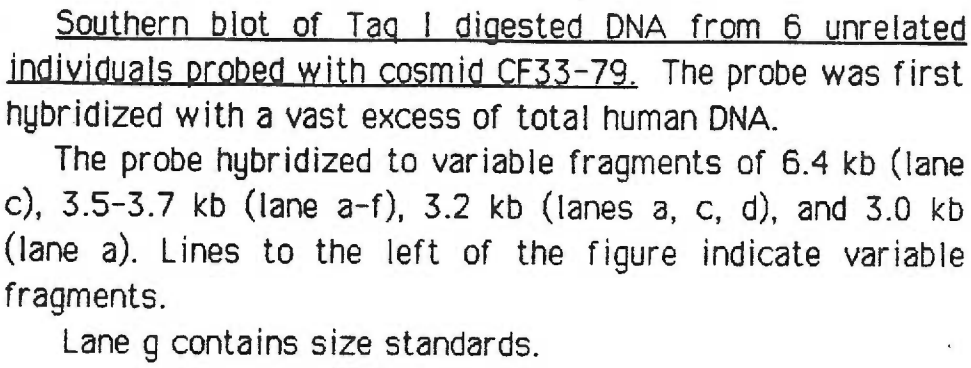


FIGURE 8



Southern blot of Taq I digested DNA from 6 unrelated individuals probed with cosmid CF33-79. The probe was first hybridized with a vast excess of total human DNA.

The probe hybridized to variable fragments of 6.4 kb (lane c), 3.5-3.7 kb (lane a-f), 3.2 kb (lanes a, c, d), and 3.0 kb (lane a). Lines to the left of the figure indicate variable fragments.

Lane g contains size standards.

CF33-79

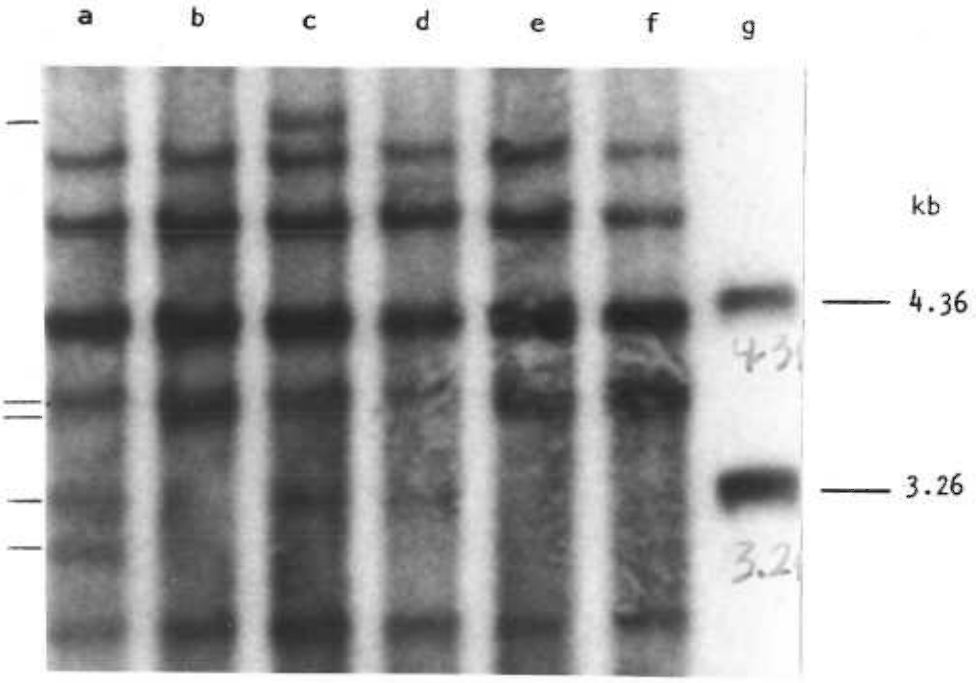
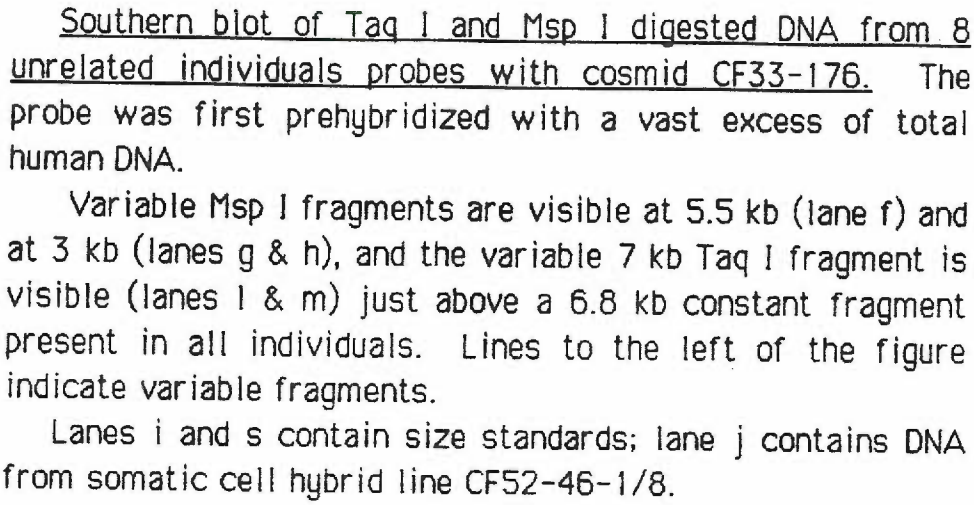


FIGURE 9



Southern blot of Taq I and Msp I digested DNA from 8 unrelated individuals probes with cosmid CF33-176. The probe was first prehybridized with a vast excess of total human DNA.

Variable Msp I fragments are visible at 5.5 kb (lane f) and at 3 kb (lanes g & h), and the variable 7 kb Taq I fragment is visible (lanes l & m) just above a 6.8 kb constant fragment present in all individuals. Lines to the left of the figure indicate variable fragments.

Lanes i and s contain size standards; lane j contains DNA from somatic cell hybrid line CF52-46-1/8.

CF33-176

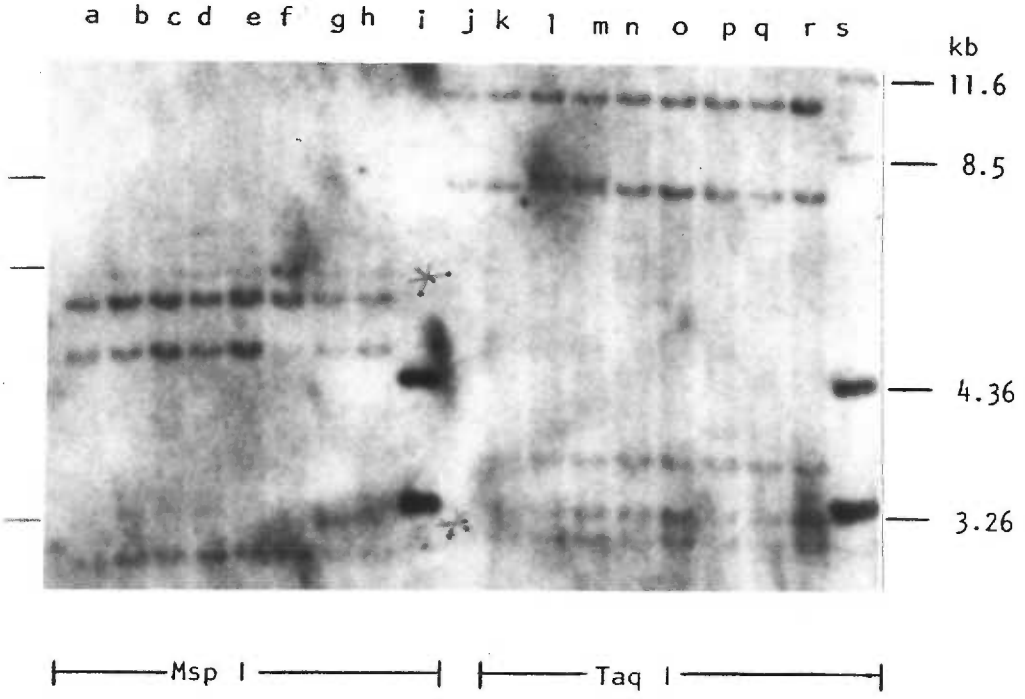


FIGURE 10

Southern blots of Rsa I and Taq I digested DNA from unrelated individuals probed with a 7.5 kb fragment of cosmid CF33-79. The fragment was first prehybridized with a vast excess of total human DNA.

The 2.5 kb Rsa I fragment and the 5.5 kb Taq I fragment seen in all individuals are constant (nonpolymorphic) fragments not seen in subsequent hybridizations using the cosmid subclones that reveal this RFLP as probes on these same blots.

For comparison of Rsa I and Taq I genotypes: lanes b & n contain DNA from the same individual who is heterozygous for the 6.2 & 3 kb Rsa I fragments and the 6.4 & 3.2 kb Taq I fragments. Lanes e & o, f & p, and g & l contain DNA from the same three individuals, respectively, all of whom are homozygous for the Rsa I 3.5 kb fragment and the Taq I 3.7 kb fragment.

Lanes a and h contain size standards. Lane i contains DNA from somatic cell hybrid line CF52-46-1/8.

CF33-79

7.5 kb fragment

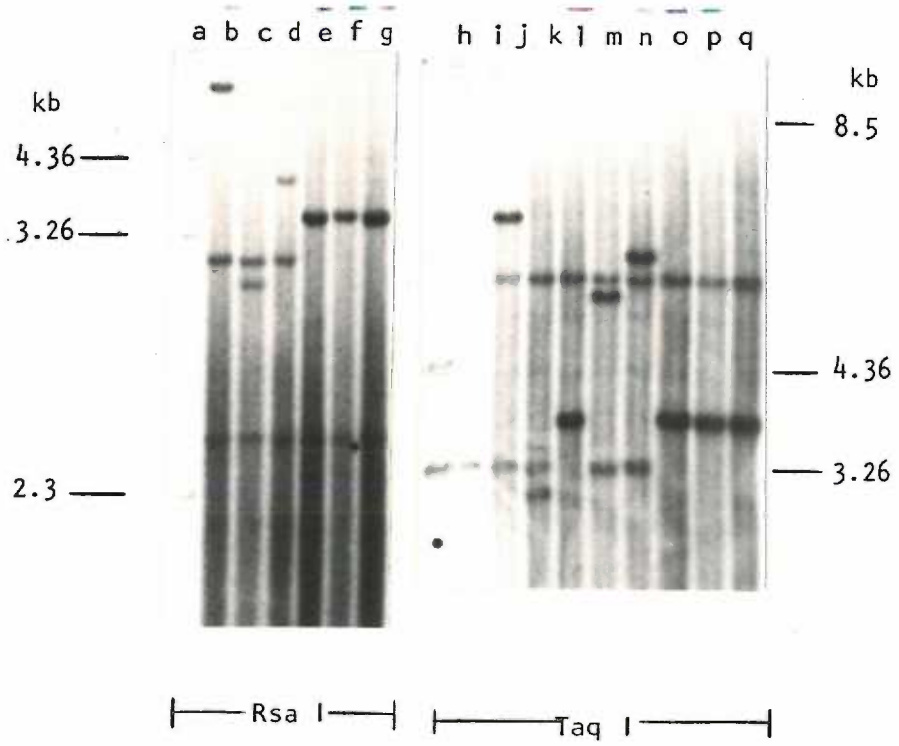


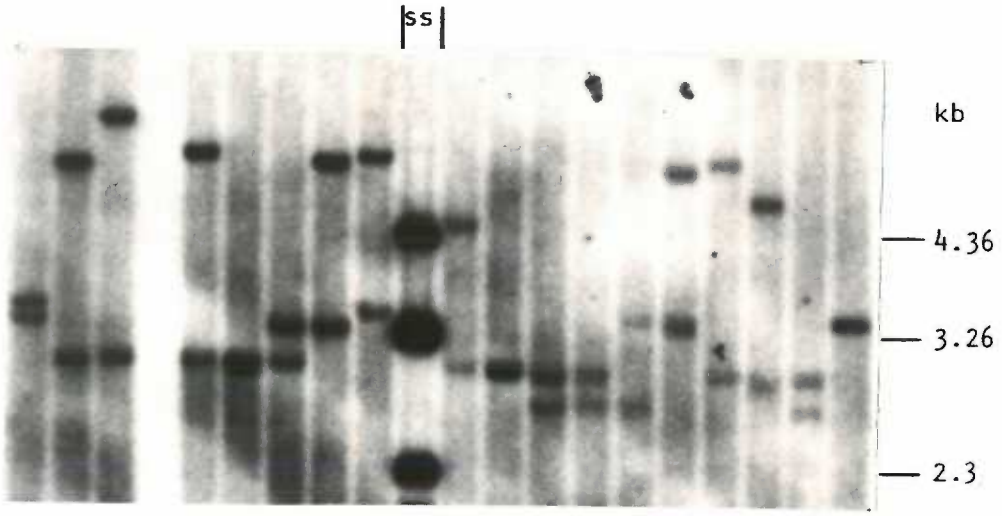
FIGURE 11

Southern blots of Rsa I digested DNA from 18 unrelated individuals probed with cosmid CF33-79 subclone p79-2-23.

Fifteen of the 18 individuals are heterozygous at this locus. Fourteen different genotypes are visible on this blot.

ss = size standards

p79-2-23



Rsa I

FIGURE 12

Southern blots of restriction digested DNA from two families probed with subclone p79-2-23.

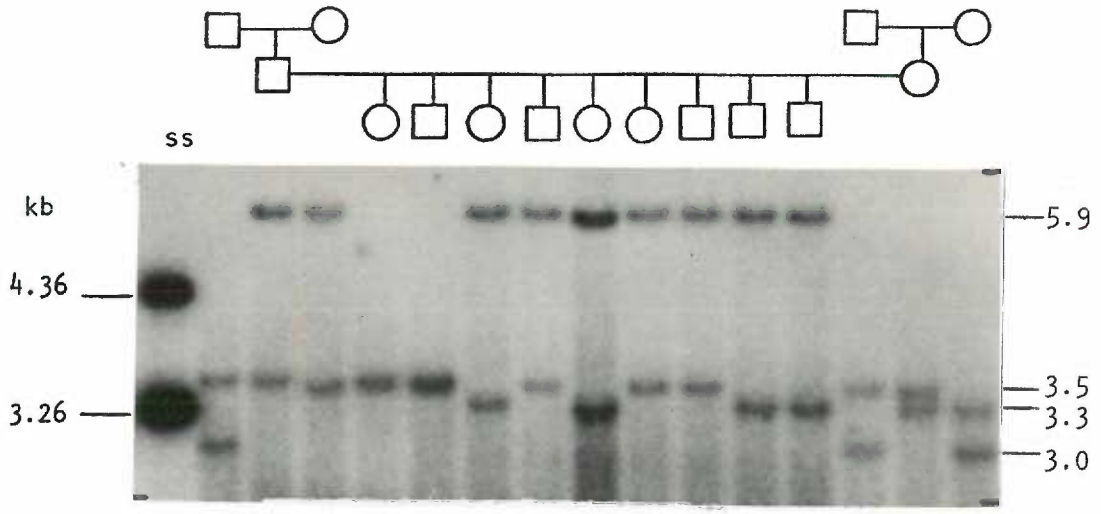
a. DNA from family K1331 dsigested with Rsa I. The father inherited a 3.5 kb fragment from his father and a 5.9 kb fragment from his mother. The mother inherited a 3.5 kb fragment from her father and a 3.3 kb fragment from her mother. The children show 3 of 4 possible genotypes: two are homozygous for the 3.5 kb fragment, four are heterozygous for 5.9 & 3.3 kb fragments, and three are heterozygous for 5.9 & 3.5 kb. fragments. No child has the heterozygous 3.5/3.3 kb fragment genotype.

b. DNA from family K1340 digested with Taq I. The father inherited a 3.7 kb fragment from his father and a 3.2 kb fragment from his mother. The mother inherited a 5.0 kb fragment from her father and a 3.2 kb fragment from her mother. The children show 3 of 4 possible genotypes: 4 are homozygous for the 3.2 kb fragment, one is heterozygous for the 3.7 and 3.2 kb fragments, and one is heterozygous for the 5.0 and 3.2 kb fragments. No child has the 5.0/3.7 kb genotype.

ss = size standards

a

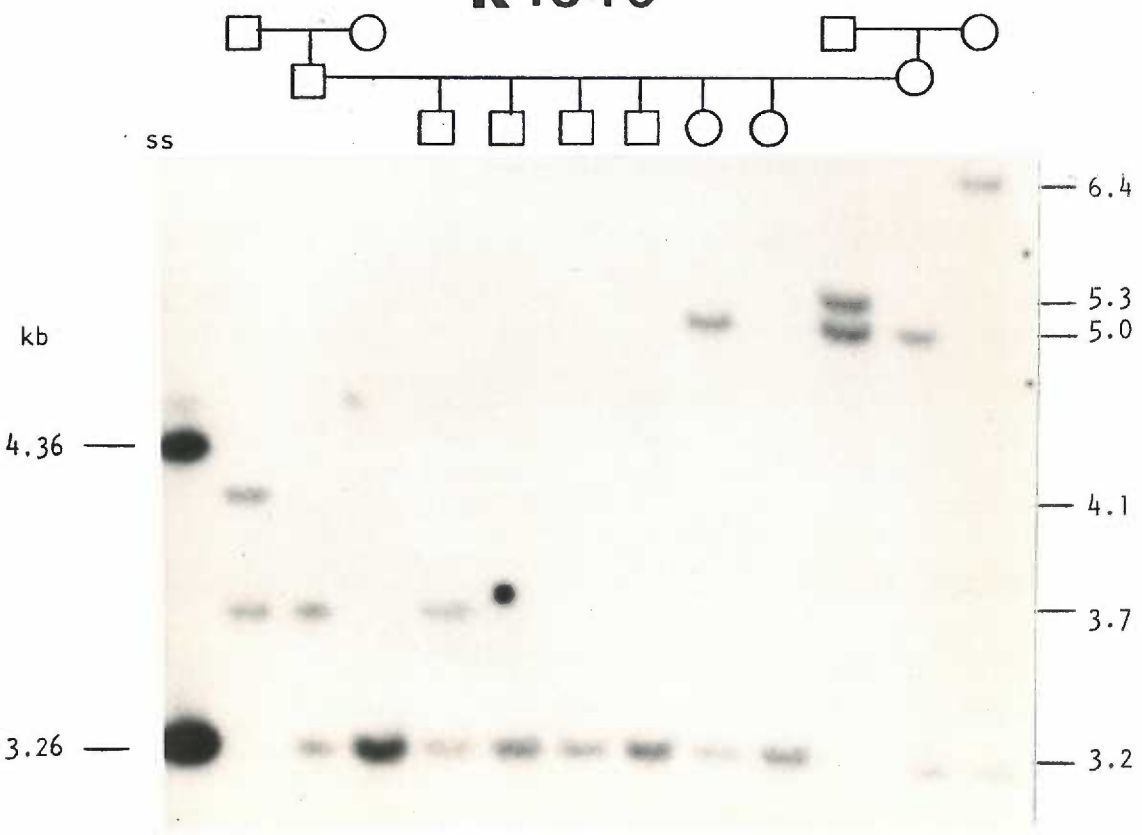
K1331



Rsa I

b

K1340



Taq I

FIGURE 13

Histogram of chromosomal distribution of silver grains from in situ hybridization of CF33-79 and p79-2-23 to chromosomes of a normal male.

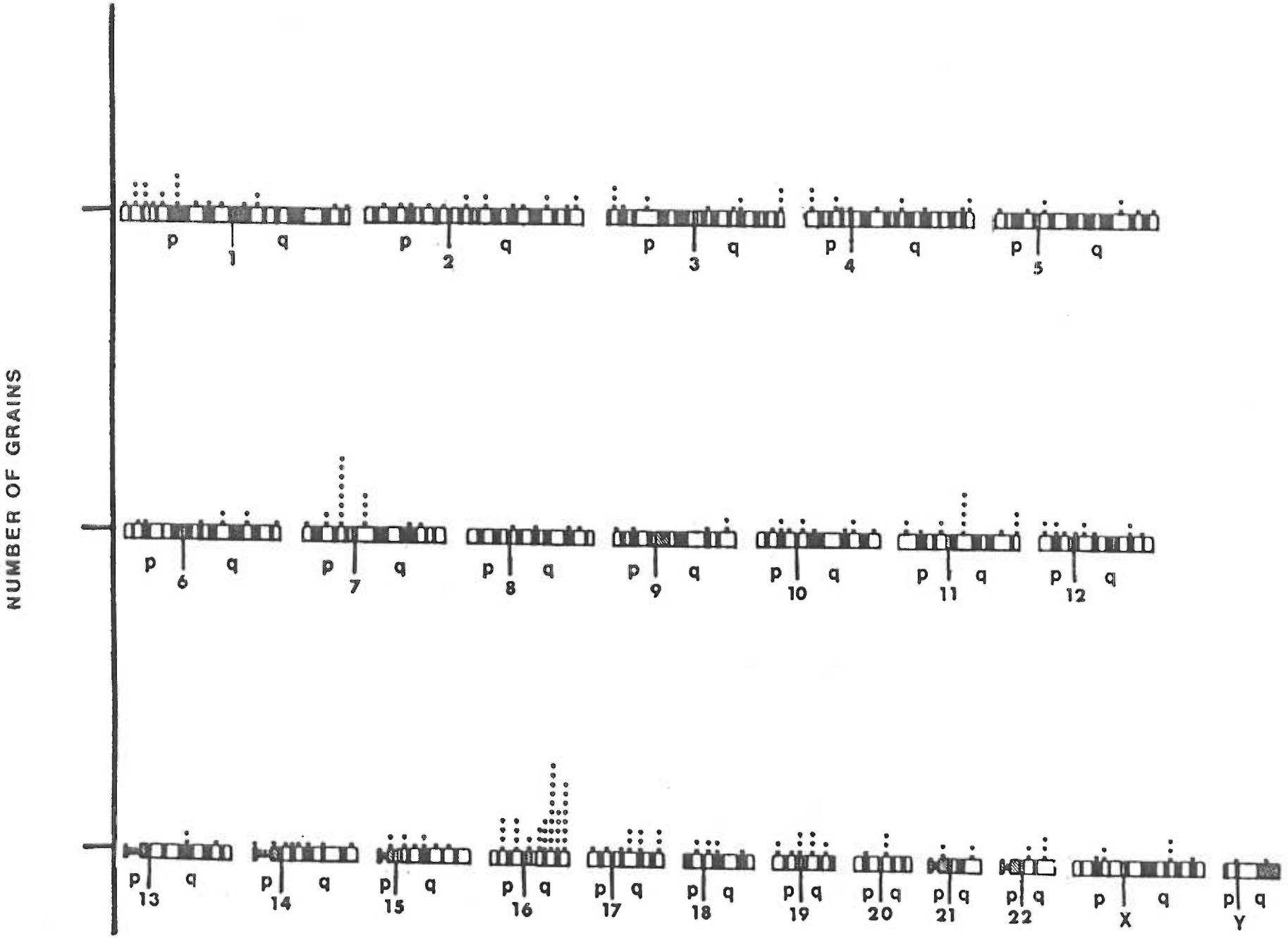

NUMBER OF GRAINS
p 1 q
p 2 q
p 3 q
p 4 q
p 5 q
p 6 q
p 7 q
p 8 q
p 9 q
p 10 q
p 11 q
p 12 q
p 13 q
p 14 q
p 15 q
p 16 q
p 17 q
p 18 q
p 19 q
p 20 q
p 21 q
p 22 q
p X q
p Y q

FIGURE 14

Southern blot of Taq I digested DNA from unrelated individuals probed with cosmid CF32-391 subclone p391-2-42.

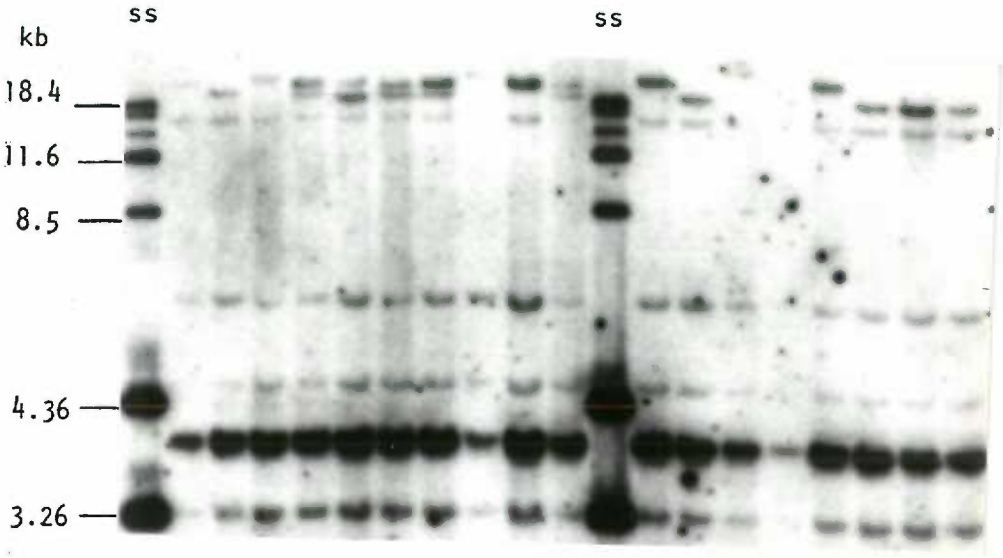
a. 18 unrelated individuals. Variable fragments are visible above 18 kb. Constant fragments are visible at 14.5, 6.2, 4.6, 3.9, and 3.3 kb.

b. Family K1340. Fragments appear to be Mendelianly inherited.

ss = size standards

p391-2-42

a



b

K1340

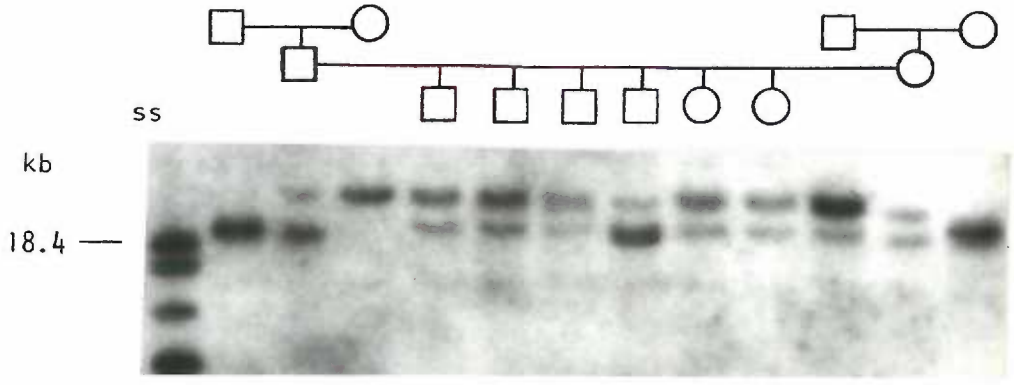


FIGURE 15

Southern blot of Taq I digested DNA from parental and hybrid cell lines probed with the insert of subclone p391-2-42. Although barely visible in some lanes, the probe hybridized to the high MW fragments in all cell lines in which it hybridized to the 3.9 and 3.3 kb fragments indicating that, if multiple loci exist, they are on chromosome 16.

ss = size standards

p391-2-42i

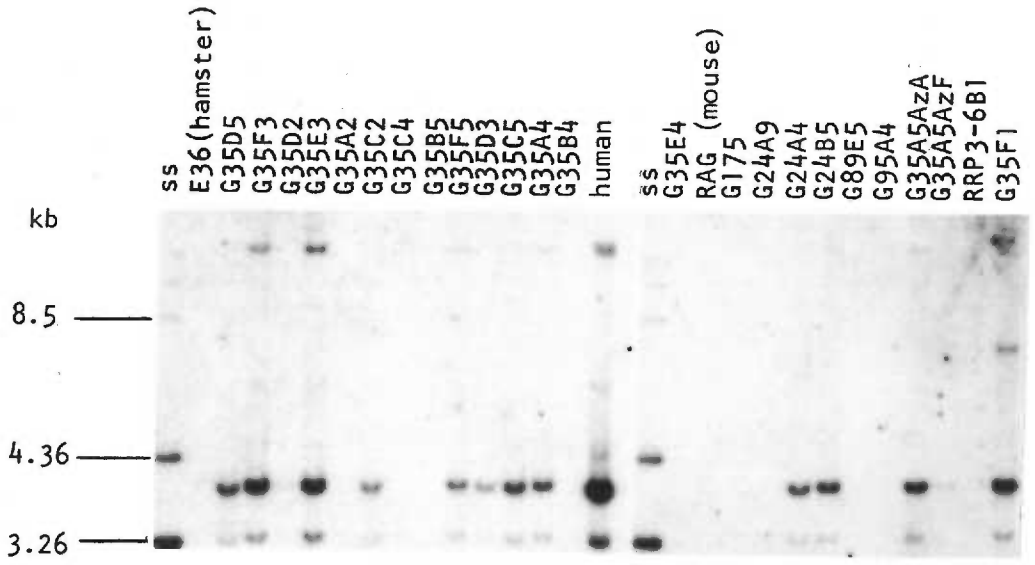

ss
E36(hamster)
G35D5
G35F3
G35D2
G35E3
G35A2
G35C2
G35C4
G35B5
G35F5
G35D3
G35C5
G35A4
G35B4
human
ss
G35E4
RAG (mouse)
G175
G24A9
G24A4
G24B5
G89E5
G95A4
G35A5AzA
G35A5AzF
RRP3-6B1
G35F1
kb
8.5
4.36
3.26

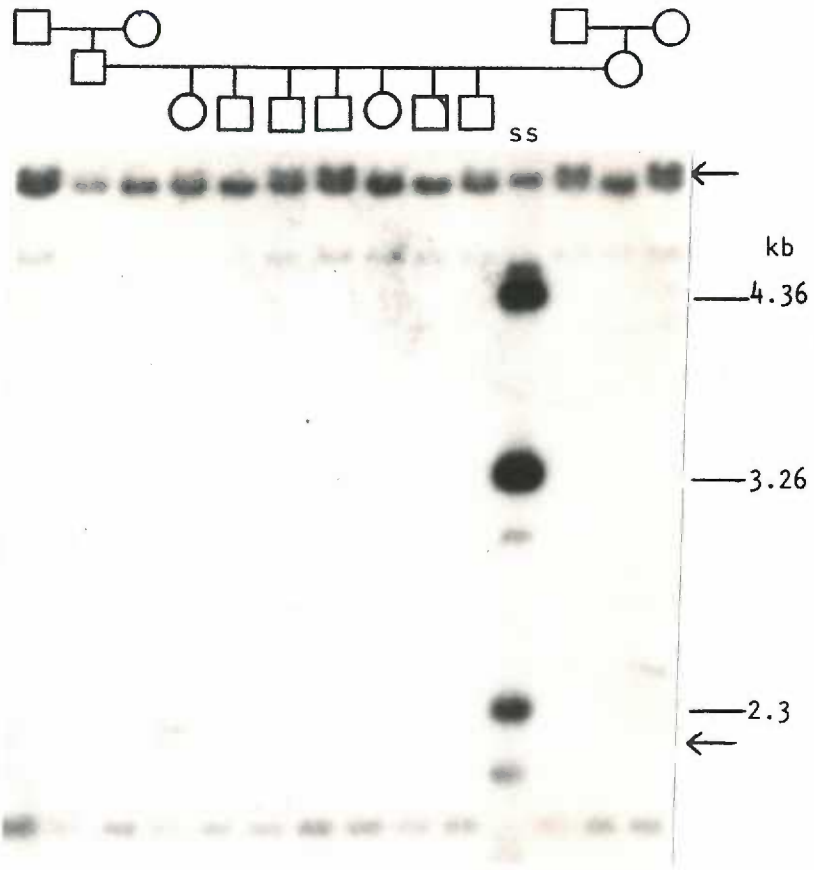
FIGURE 16

Southern blot of Msp I digested DNA from family K1345 probed with cosmid CF33-176 subclone p176-1-108. Arrows point to the 5.6 kb and 2.2 kb fragments that do not conform to Mendelian inheritance, i.e. the 5.6 kb fragment is present in 3 grandparents and several children; the 2.2 kb fragment is present in one child but does not appear in either parent or grandparents.

ss = size standards

p176-1-108

K 1345



C. Discussion

A partial cosmid library specific for human chromosomes 11q and 16q was constructed and screened for RFLPs.

Construction of the library took several months and was fraught with technical difficulties. Other laboratories have reported similar experiences, and statements such as "every success has been preceded by 2 or 3 failures" (P. Little, pers. comm., 1986) were encountered throughout the course of this research.

The biggest problem was getting ligation to work. Various ligation mixes and conditions were used, and a Bam HI phosphatased Kos I vector was tried before use of the pJB8 arms. No single cause for the difficulties was found but rather it seemed to be a matter of optimizing all conditions. For example, the control λ ligation worked every time whether or not the test ligation worked so that the enzyme or reagents did not appear to be the problem. Kos I was abandoned in favor of pJB8 because a good phosphatased preparation was not achieved.

Ligation finally did work, but the number of recombinant clones obtained per µg of eukaryotic DNA after *in vitro* packaging and transduction into bacterial hosts was many times lower (5500/µg) than most reports of 300,000-500,000 clones/µg (Collins & Hohn, 1978; Ish-Horowicz & Burke, 1981; Dillela & Woo, 1985; P. Little, pers. comm., 1986) although lower efficiencies of 10,000-100,000

have been reported (Little & Cross, 1985), and many papers do not report the efficiency.

The factor most often cited as a critical determinant in successful cosmid library construction is the quality of the starting genomic DNA (Dillela & Woo, 1985; P. Little, pers. comm., 1986). It is important to start out with very high molecular weight DNA (>150-200 kb) so that after partial digestion, a majority of the molecules will have restriction sites at both ends. According to P. Little (pers. comm, 1986), a partial digest fraction of 40-50 kb isolated from DNA of starting size 100 kb has only 36% of its molecules with restriction sites at both ends. The broken molecules compete in the ligation reaction and greatly decrease the efficiency of ligation.

Even if the starting DNA is of good size, such as in my library, problems can arise as a result of DNA degradation at subsequent steps. We were repeatedly unable to get sharp bands of DNA during size fractionation but rather ended up with smears over a wider size range (See figure 3, "before ligation"). Fractions that had the bulk of DNA in the 35-50 kb size range were pooled. In retrospect, these smears probably resulted from DNA degradation while on the gradient (P. Little, pers. comm., 1986). Such degradation would increase the number of damaged ends thereby decreasing the efficiency of ligation. This degradation probably accounts for the low numbers of clones obtained per µg of ligated genomic DNA in my library.

Poor packaging mixtures have also been cited as major culprits in cosmid library failures (Dillela & Woo, 1985), however, the Packagene used for this library worked well when used by me and by others for other projects in the laboratory.

As previously stated, efficiency was not critical as a complete library was not the goal. Based on previous results from screening a random library in this laboratory, 3000 clones from the somatic cell hybrid DNA was thought to be a sufficient number for RFLP searches of 11q and 16q.

In this laboratory, a total of 51 random cosmids from a library made by Ed Fritsch have been screened over a 2-year period. When used as probes on Southern blots, 42 gave autoradiographs on which bands could be seen; 14 of these, or 27%, displayed multiple RFLPs on Taq I and/or Msp I blots and were selected for further study.

We estimated that a somatic cell hybrid DNA library of 3000-5000 clones would yield about 100 human clones. If only 15% of these clones revealed multiple polymorphisms, a conservative figure bases on results with the random library, this would give about 15 multiple RFLPs spread between chromosomes 11q and 16q.

The initial cosmid colony screening with total human and total mouse DNA probes yielded fewer putative human clones than expected, 50 vs ca. 85, based on the reported chromosomal composition of the somatic cell hybrid line used to construct the library. Of 43 of the 50 clones further investigated, only 20 showed

definite hybridization when used as probes on Southern blots of total human DNAs from several individuals for a final figure of about 0.7%, much lower than the 2.5% expected. A phage library has recently been constructed in this laboratory by C. Maslen-McClure using the same somatic cell hybrid DNA used to construct the cosmid library. Preliminary results indicate that the number of putative human DNA-containing clones observed by colony hybridization with a total human DNA probe is about 1% (C. Maslen-McClure, pers. comm., 1986) which is compatible with my findings.

The screening method was obviously far from perfect as shown partly by the fact that 6 clones initially picked as human because they did not hybridize to mouse DNA later showed strong hybridization to the somatic cell hybrid DNA and none to total human DNAs on the same Southern blots. Many clones did not hybridize to either total human or total mouse (approximately 20% based on counts from plates 32, 33, &34); some of these probably represented clones with poor growth and/or transfer to filters. Some may have been clones containing too few repeats for the stringency of hybridization used. Hybridization at varying stringencies may have resulted in the detection of more clones containing human inserts.

Actually, various hybridization conditions were tried. Hybridization conditions were initially determined by probing control colony filters containing known human cosmids with nick-translated total mouse and total human DNA. The human probe hybridized to

most of the human clones; the mouse DNA did not hybridize to any of the human clones. When these conditions were used for the cosmid library colony filters, however, both human and mouse probes hybridized to most of the colonies indicating lack of species specificity and/or non-specific colony background interference..

Therefore, early in the colony screening stage, two major changes were made in the protocol: 1) a switch from nick-translated probes to oligolabeled probes and 2) the elimination of dextran sulfate from the prehybridization/hybridization solution. The protocol (as described in Materials and Methods) was from D. Housman's laboratory where it is used to pick out human DNA-containing clones from mouse-human somatic cell hybrid phage libraries (B. Handelin, pers. comm., 1985). The stringency is lower than that used with nick-translated probes on the control colony filters described above; however, this protocol was not tested using the control filters.

It is also possible that some clones may have deleted part or all of the eukaryotic DNA during the interval of transfer from the original plates to storage wells or gridded plate and then to replica filters. Cosmid instability has been noted (Collins & Hohn, 1978; Grosveld et al, 1981; Gumucio et al, 1985; P. Little, pers. comm., 1986). However, cosmids are usually most stable on solid media (Gumucio et al, 1985).

About 10-15% of clones hybridized to both total human and total mouse probes. Some of these probably represented clones containing noncontiguous pieces of DNA, one mouse and one human. The genomic DNA, although size fractionated, was not phosphatased prior to ligation, and fragments smaller than 20 kb may have been present in the pooled fractions. Such non-contiguous DNA has been observed in other cosmid libraries (Grosveld et al, 1981; M. Litt, pers. comm., 1985). Another possibility is that some clones may have hybridized to both total human and total mouse because of stringency conditions so that once again, varying hybridization conditions may have sorted out some of these clones.

Five of the 43 clones for which cosmid DNA was isolated, nearly 12%, had obvious deletions. In addition, 2 of the 3 clones being grown up repeatedly for further investigation of possible RFLPs were found to delete under large scale (175-350ml) liquid culture conditions. Transduction of the repackaged intact cosmid DNA into a more stringent Rec A^- strain, DK-1, did not alleviate the problem.

The molecular basis for the deletions is not known (Gumucio et al, 1985), but Little and Cross (1985) speculate that, as Col E1-derived cosmids such as pJB8 appear to have an inverse relationship between size and copy number, small molecules generated by illegitimate recombination, by virtue of their increased copy number, will eventually dominate a population. They have developed a new cosmid vector, based upon the phage λ origin of

replication, that shows a higher and more constant copy number and correspondingly increased stability. P. O'Connell(pers. comm., 1986) in R. White's laboratory has found that doing DNA preparations from cultures before they reach saturation appears to decrease the number of observed deletions.

Another major problem with this cosmid library was the consistently poor yields of cosmid DNA from both minipreps and large scale preparations using CsCl gradients. One would expect lower yields than for plasmid preps due to the increased size and decreased copy number (Little & Cross, 1985), but the repeated failure to get cosmid DNA banding in CsCl gradients, when plasmid preparations done simultaneously worked well, remains unexplained. Using the same host/vector system, Ish-Horowicz & Burke (1981) reported yields up to 10 mg/L using similar large scale preparative methods; most other papers reviewed and referenced above did not report the cosmid DNA yields.

Of the 20 cosmids that hybridized to human DNAs on Southern blots, 8 revealed possible RFLPs of which 2 (3 if the duplicate to CF33-79, CF32-435, is considered) were fully characterized and found to be highly polymorphic loci on chromosome 16q. These figures, considering only the 20 cosmids proven to contain human DNA inserts, are more in line with expectations based on results from the random library. If all 8 possible RFLPs had been mapped using the somatic cell hybrid mapping panel, it is probable that

several would have mapped to chromosome 11q. In fact, CF33-176, from which the Msp I "RFLP" was found to be an artifact and from which no single copy probe showing the Taq I polymorphism was found, mapped to 11q when used to probe the mapping panel.

In summary, the low number of human clones found was probably due to a combination of factors including colony screening techniques, deletions, and an inability to get good cosmid DNA preparations. The reasons for theses difficulties remain obscure.

The Human Polymorphism Study Center (CEPH) has recently been organized under the aegis of Jean Dausset at the College de France in Paris; its primary goal is to produce a detailed map of genetic markers encompassing all the chromosomes of the human genome (Marx, 1985). Such a map will greatly facilitate the localization and study of disease-causing genes. Toward realization of this goal, the Center has collected, from investigators worldwide, cell lines of 40 large multigenerational families. DNA from these families will be provided to investigators possessing DNA probes revealing RFLPs if they agree to determine the inheritance pattern in all informative "CEPH" families. These data will then be provided to CEPH so that comparisons can be made with other RFLP markers to establish linkage patterns.

As of HGM 8, two compound polymorphic loci had been mapped to chromosome 16q. Adenine phosphoribosyltransferase (APRT) has two Taq I RFLPs, with PICs of 0.27 and 0.28, that map to 16q22. The other locus is haptoglobin, at 16q22.1, where three RFLPS with PICs of 0.25, 0.33, and 0.37 have been located (Willard et al, 1985). Published data do not allow evaluation of the heterozygosities or PICs of either of these compound loci. Other genes assigned to 16q include diaphorase-4 (16q12->q22), glutamic-oxaloacetic transaminase (16q12-q22), lecithin-cholestrol acyltransferase (16q22), and the metallothionein gene family (16q22); of these 4 genes, DNA probes are available only for metallothionein. A fragile site exists at 16q22.1 (Willard et al, 1985).

The CF33-79 RFLP characterized in this study has 80% heterozygosity and a calculated PIC value of 0.77; it should prove quite useful in constructing a genetic linkage map of 16q. Preliminary studies with CF32-391 suggest that it may also be useful. More precise localization of these RFLPs by in situ hybridization to metaphase chromosomes of individuals with various translocations and deletions of 16q is underway in E. Magenis' laboratory.

These probes as well as others from the phage library currently being screened by C. Maslen-McClure will be given to R. White who has agreed to use them in linkage studies with the CEPH families to construct linkage maps of 11q and 16q. Such maps will constitute a

part of the total human genomic linkage map being constructed by R. White and others as part of the CEPH project. Probes will also be made freely available to interested investigators for use in linkage studies with various hereditary disease loci.

III

Part 2: Manuscript

Four Restriction Fragment Length Polymorphisms Revealed by Probes from a Single Cosmid Map to Chromosome 19

L. Bufton[1], G.A.P. Bruns[2], R.E. Magenis[1,3], D.Tomar[1,3], D. Shaw[4], D. Brook[4], and M. Litt[1,5]

American Journal of Human Genetics
Vol 38 (1986)

Received July 5, 1985; revised August 30, 1985.
This work was supported by grants R01-GM32500 and P01-HD18658 from the National Institutes of Health. L.B. was supported by a Life and Health Insurance Medical Research Fund Scholarship.

[1]Dept. of Medical Genetics, Oregon Health Sciences University, Portland, OR. 97201.
[2]Genetics Division and Mental Retardation Research Center, Children's Hospital, Boston, MA 02115.
[3]Crippled Children's Division, Oregon Health Sciences University.
[4]University of Wales College of Medicine, Cardiff CF4 4XN, Wales.
[5]Dept. of Biochemistry, Oregon Health Sciences University.

SUMMARY

We have discovered and characterized a compound polymorphic locus on chromosome 19, defined by an arbitrary genomic DNA segment cloned into a cosmid vector. Four different restriction fragment length polymorphisms with minor allele frequencies equal to or greater than 10% are revealed by Southern hybridization of subclones of cosmid 1-13 with Taq I, Msp I, Bam HI and Hind III digests of human DNAs. Seventy two percent of unrelated individuals are heterozygous at one or more loci, and 7 of the 24 possible haplotypes occur with frequencies of 3-38%. Using a somatic cell hybrid panel, we have mapped this locus to 19p13.2→19q13.3, whereas in situ hybridization suggests the probe is on 19p. Taken together, these results suggest localization to 19p13.2→19cen. The locus revealed by probes from cosmid 1-13 has been designated D19S11.

INTRODUCTION

Restriction fragment length polymorphisms (RFLPs) are useful markers for gene mapping [1]. More than 160 RFLPs in the human genome have been reported, and the number is expanding rapidly [2]. Most of these RFLPs, however, have only 2 alleles with maximum possible heterozygosities of 50%, and in practice, heterozygosities are usually considerably smaller. Since frequent heterozygosity is a critical requirement for linkage studies, we have attempted to use cosmids from a human genomic library as sources of probes. Cosmids have much larger human DNA inserts (35-45 Kb) than phage libraries which are most often used in similar studies. In principle, the longer the segment, the more sites can be examined, increasing the possible number of polymorphisms per segment. For example, if 3 polymorphic sites were found in one DNA segment and the eight possible haplotypes occurred with equal frequency, the compound locus would be a highly informative marker with 87.5% heterozygosity. The usefulness of actual compound loci will depend on allele frequencies and on the degree of linkage disequilibrium, which could be substantial over these short distances. However, we [3] and others [4-6] have observed that the degree of linkage disequilibrium between closely linked RFLPs may often be small.

A novel blot hybridization method, using an entire cosmid to probe Southern transfers of restriction digests of genomic DNA, greatly increases the usefulness of this approach [3]. By prehybridizing the probe with a vast excess of total human DNA, sequences in the probe

that are homologous to genomic repeats are rendered unable to hybridize with the Southern transfers, and therefore allow visualization of single and low-copy bands.

Results previously reported by us have confirmed the usefulness of this method. Probes from one cosmid, mapped to chromosome 2q, revealed 7 polymorphic loci with 93% of unrelated individuals heterozygous at one or more loci [3]. Probes from a second cosmid mapping to 12q revealed a compound locus of 3 polymorphic loci with 74% of unrelated individuals heterozygous at one or more loci [7]. Here, we describe a compound polymorphic locus on chromosome 19. Seventy two percent of unrelated individuals were heterozygous at one or more of 4 closely linked loci revealed by probes derived from cosmid 1-13, making this a very useful marker for linkage studies.

MATERIAL AND METHODS

Human DNA isolation, cosmid preparation and screening, subcloning, Southern transfer, and hybridization procedures were described previously [3]. The cosmid characterized in this study was from a partial human genomic DNA library obtained from Ed Fritsch, The Genetics Institute, Boston. Linkage studies were done using DNAs isolated from lymphoblast cell lines of large 3 generation Utah pedigrees as previously described [3]. The cell lines were obtained from the Human Genetic Mutant Cell Repository, Institute for Medical Research, Camden, N.J. In all informative pedigrees tested for linkage, determination of grandparental genotypes revealed the linkage phase

in the parents, allowing recombinant and non-recombinant children to be identified by examination of their genotypes.. LOD scores were calculated at a recombination fraction of zero.

Parental and Hybrid Cell Lines: For initial experiments, 23 cell lines were used as a mapping panel. The G35 cell lines are human-hamster hybrids derived from fusion of the Chinese hamster cell line E-36 with white blood cells (WBC) from a female carrier of the X/19W translocation t(X;19)(q23-25::q13)[8]. The G17 and G24 cell lines are human-mouse hybrids derived from fusion of the mouse cell line RAG with WBC of the X/19W translocation carrier (G17 lines) or the X/19B translocation t(X;19)(q1::p13)[9]. The G89E5 and G95A4 cell lines were derived from fusions of WBC from a karyotypically normal male with the Chinese hamster cell lines YH21 and Wg3h, respectively. G35A5AzA and G35A5AzF are subclones of G35A5 selected in 8-azaguanine for loss of the der19 translocation chromosome. These hybrid cell lines were characterized by both isozyme and cytogenetic techniques [10,11]. In addition, DNAs from these hybrids have been analysed with cloned DNA probes for all chromosomes except the Y.

We also used a mapping panel consisting of three human x mouse hybrid cell lines containing chromosomes derived from translocations involving chromosomes X and 19 as follows: line G175AoXiB, der19,t(X;19)(Xqter→Xq23/25::19q13→19pter); line G24B2AMB, der19,t(X;19)(Xqter→Xq11/13::19p13→19qter); line GM89A99c7B, derX,t(X;19)(Xpter→Xq22::19q133→19qter.) [9].

In Situ Hybridization. Subclone 1-82 was labeled by nick-translation according to the method of Harper and Saunders [12] to a specific activity of 4×10^7 dpm/µg using [^{3}H] TTP (65 Ci/mmol) and [^{3}H] dCTP (60 Ci/mmol, Amersham, Arlington Heights, Ill). In situ hybridization to metaphase spreads from normal male cells was performed according to Harper and Saunders [12]. Chromosome preparations were treated for 1 hr with RNAse (100µg/ml) in 2X SSC at 37^0 and then denatured in 70% formamide at 70^{0}C. Probe 1-82 was diluted to a final concentration of 50 ng/ml in pH 7.0 hybridization buffer that contained 50% formamide, 2X SSCP, 10% dextran sulfate, and 50µg/ml salmon sperm DNA. Denatured chromosomes were hybridized for 12 hr at 37^{0}C; excess probe was then removed by washing in three changes of 2X SSC, 50% formamide at pH 7.0 for 3 min each, followed by 5 rinses, for 2 minutes each, in 2X SSC. All washes were at 39^{0}C, two degrees above the hybridization temperature. Following dehydration with an alcohol series, Kodak NTB-2 liquid emulsion was applied to the slides for a 5-day exposure. Slides were developed and were R-banded using a modification of the technique of Schweizer [13)] For photographs illustrating grain distribution, preparations were destained and restained with Wright's stain.

RESULTS

Initially, the entire cosmid 1-13 was used, after prehybridization with a vast excess of total human DNA, to probe Southern transfers of DNAs from six to nine unrelated individuals digested with 10 different restriction enzymes. This initial screening revealed variable bands with several enzymes, including Taq I, Msp I (fig. 1), Hind III and Bam HI (not shown.) Thus cosmid 1-13 revealed polymorphisms involving Msp I fragments of 4.3, 3.1 and 2.6 kb, Taq I fragments of 4.3 and 2.3 kb and a Hind III fragment of 5.8 kb. The cosmid was digested separately with these enzymes, and, where present, fragments corresponding in size to the polymorphic fragments in genomic digests were isolated from agarose gels and used as probes of colony filters containing Sau 3A subclones of the cosmid. Subclones that gave signals with a cosmid restriction fragment probe but not with total human DNA were used as probes against appropriate restriction digests of human genomic DNA from unrelated individuals.

Identification of Fragments Revealing Polymorphic Restrictior Endonuclease Sites. Two subclones, 1-82 and 2-21, revealed a total of 4 polymorphic loci, α, β, γ and δ. Table 1 summarizes important characteristics of these polymorphic loci; blots illustrating some of these loci are shown in Figure 2.

Subclone 1-82 revealed RFLPs with Taq I (fig. 2A) , Msp I, Hind III and Bam HI. We believe that the Taq I and Msp I polymorphisms

revealed by 1-82 (the α locus) reflect the insertion-deletion of approximately 100 bp within the region detected by this probe. This is supported by the exact correlation between Taq I and Msp I genotypes in 33 unrelated individuals typed at both loci and by the equivalent 0.1 kb size difference between allelic fragments found with both Taq I and Msp I. The α locus is also revealed by a third subclone, 2-11 (fig. 2B). A model for the α locus consistent with this data is shown in figure 2C.

The δ locus, revealed with subclones 1-82 and 2-11, gives allelic fragments of 6.2 and 5.4 kb with Hind III (fig. 2D) and most likely consists of a polymorphic central restriction site flanked by non-polymorphic sites with the probe hybridizing to only one (5.4 kb) of the 2 fragments (5.4 & 0.8 kb) created when the central Hind III site is present.

Subclone 1-82 also revealed a 3-allele Bam HI polymorphism, with variable fragments of 14.0, 11.2 and 10.5 kb (data not shown.) This polymorphism is best explained by a model involving two neighboring polymorphic sites. An insertion/deletion model is unlikely because the RFLPs seen with other enzymes have differences in allelic fragment sizes that do not match those seen with the BamHI RFLP. The 3-allele Bam HI polymorphism is also revealed by subclone 2-21.

The β locus is revealed when subclone 2-21 is used to probe Southern transfers of Msp I digested DNA (fig. 2E). The probe hybridizes to 2 allelic fragments of 2.54 and 2.47 Kb. As with the δ locus, the genome for the β locus most likely consists of 2

nonpolymorphic Msp I sites flanking a central polymorphic site with the probe hybridizing to one of the 2 fragments created when the polymorphic site is present.

The polymorphic loci were shown not to be artifacts due to partial digestion by digesting one or more genomic DNAs containing the larger allelic fragment with increasing amounts of restriction enzyme and observing the persistence of this fragment in the presence of excess enzyme. In addition, the observation of Mendelian inheritance of the various alleles in our family studies rules out sporadic partial digestion or DNA contamination as possible artifacts.

The isolation of probes 1-82 and 2-21 from the same cosmid suggests, but does not prove, that the polymorphic loci revealed by these probes are closely linked. Because of the occasional presence in cosmid libraries of clones containing inserts derived from non-contiguous regions of genomic DNA [3], the assumption of close linkage of loci revealed by subclones from a single cosmid must be verified. In the case of cosmid 1-13, close linkage of subclones 1-82 and 2-21 is supported by Southern blotting and by family studies using these two probes.

Southern Blotting. Subclones 1-82 and 2-21 both hybridized to a 12.5 kb Bam HI fragment of cosmid 1-13 (data not shown). When these same probes were hybridized separately to Southern transfers of BamHI digests of human genomic DNA, both probes revealed 11.2 kb fragments (fig. 3). These findings are consistent with other evidence, not shown, that the 12.5 kb Bam HI fragment is at one end

of the cosmid insert, includes a portion of vector DNA, and lacks the next genomic Bam HI site. Hybridization of both subclones to genomic Bam HI fragments of identical sizes strongly supports colinearity within the genome of the region of the cosmid containing the subclones.

Linkage analysis. Linkage studies were done using 5 large Utah families, each with 4 living grandparents and 5-8 children. Linkage phase was known for all parents. LOD scores for linkage at a recombination fraction of zero were 2.7 for the α-β pair, 5.7 for β-γ, and 0.9 for α-γ (only 3 children were informative for the α-γ pair). No recombinants were observed. Linkage studies involving the δ locus were not done.

Somatic cell hybrid panel. Eco RI digests of DNAs from parental mouse and hamster cells and from 23 human-rodent hybrid cell lines were probed with [^{32}P] labeled inserts from subclone 1-82 (fig. 4). Subclone 1-82 cross hybridizes weakly with the hamster parent to give a band at 4.3 kb and more strongly with many of the hybrid cell lines to give bands at 4.2 and 2.8 kb. The faint bands at ca. 4.3 kb seen in figure 4 for cell lines G89E5, G95A4, G35AzA and G35AzF are attributed to cross reaction of probe 1-82 with hamster DNA because on blots made from gels run longer , for better resolution, these bands were seen to be distinctly slower than the human-specific 4.2 kb band in cell lines such as G24B5. Also, it should be noted that cell lines G89E5, G95A4, G35AzA and G35AzF lack the 2.8 kb human specific

band. As shown in table 2, these results are completely concordant only with assignment to chromosome 19.

More precise localization was made using 3 additional somatic cell hybrid cell lines that had retained different regions of chromosome 19 (fig. 5). Probe 1-82 hybridized to cell lines G175AOXiB and G24B2AMB but not to GM 89A99c7B (fig. 6) indicating localization to the region 19p13.2→19q13.3.

In situ hybridization. A representative, sequentially stained metaphase spread from the in situ hybridization studies is shown in figure 7. The silver grains revealing hybridization of the probe are observed with standard staining, with the particular chromosomes to which the probe has hybridized identified by sequential fluorescent R-banding. Figure 8 summarizes the results of scoring 100 cells from a normal male. Eighteen of the 100 cells had a grain localized to the region 19 pter→cen; the remainder of the grains were randomly distributed over the chromosomes.

Population genetic characterization of RFLPs. The observed and expected genotype frequencies for all four loci, based on 38-62 unrelated individuals tested for each locus, were in good agreement, indicating that each locus is in Hardy-Weinberg equilibrium. Of 65 unrelated individuals tested, 47 (72%) were heterozygous at one or more loci. This is a conservative figure because 11 individuals homozygous at 2 or 3 loci were not tested at all 4 loci and because the α locus, thought to represent an insertion-deletion polymorphism, appears to have more than 2 alleles which, because of their closeness

in size, could not be consistently resolved. Furthermore, another subclone of the cosmid, p1-25, reveals two other highly polymorphic loci as well as the α locus, which may be closely linked to that locus (see below.)

The haplotypes of 34 chromosomes from 17 unrelated individuals were determined from individuals in whom only one of the 4 sites was heterozygous (table 3.) Seven of the 24 expected haplotypes were observed, with frequencies ranging from 3 to 38%. Using the observed haplotype frequencies, we calculate a polymorphism information content (PIC) [1] of 0.72. Using expected haplotype frequencies, caclulated from the allele frequencies for the 4 loci, assuming no linkage disequilibrium, we calculate a PIC value of 0.87. Although the number of complete haplotypes determined so far is too small to permit a meaningful analysis of linkage disequilibrium in this system, the difference between observed and expected PIC values suggests that such disequilibrium is significant.

Nomenclature. At the 8th International Human Gene Mapping Workshop, the Committee on Human Gene Mapping by Recombinant DNA Techniques designated the compound polymorphic locus revealed by probes from cosmid 1-13 as D19S11 [14].

DISCUSSION

We have described a set of 4 RFLPs derived from a single cosmid that constitutes a compound polymorphic marker in the human genome. The use of the entire nick-translated cosmid as a probe, after

prehybridization with a vast excess of human DNA to render probe sequences homologous to genomic repeats incapable of hybridization with Southern transfers [3], was a crucial element in the discovery of this compound marker. Close linkage of the RFLPs was verified by family studies and by Southern blotting studies. At least 72% of unrelated individuals are heterozygous at one or more loci, making this compound locus a useful marker for human gene mapping.

We have used a somatic cell hybrid panel to localize the compound locus to chromosome 19p13.2→19q13.3. Our In situ hybridization studies suggest that the locus is on 19p. Taken together, these data suggest localization to 19p13.2→cen.

In work to be published elsewhere (M. Litt, U. Surti, E. Kumlin, et al, manuscript in preparation), we describe results obtained with subclone p1-25 of cosmid 1-13. When used on Southern transfers of Taq I-digested DNA, this probe reveals the α polymorphism (table 1) plus two to four additional fragments whose sizes vary in unrelated individuals, suggesting the existence of 2 additional polymorphic loci. These additional loci appear to be highly heterozygous, with a total of at least 20 alleles observed so far. In situ hybridization of p1-25 to a metaphase spread from an individual with a 19/X translocation with a breakpoint close to the centromere localizes this probe to 19p, suggesting that the 2 additional RFLPs are closely linked to the α locus and supporting our assignment of the α, β, γ, and δ loci to 19p13.2→cen. By means of family studies as well as studies with DNA from complete hydatidiform moles homozygous at all loci [15], attempts are in progress to assign the numerous Taq I bands revealed

by p1-25 to specific loci and to measure the genetic distance between these loci and the RFLPs described in this paper. (M. Leppert, personal communication, 1985.)

Numerous genes have been mapped to chromosome 19 [16] including the gene for myotonic dystrophy, an autosomal dominant degenerative muscle disease. The cosmid 1-13 RFLPs described here are being used in linkage studies with this disease [17,18] and should also prove useful in linkage studies of other genes located on chromosome 19.

ACKNOWLEDGEMENTS

We thank Ruth Litt for expert technical assistance.

REFERENCES

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

2. Skolnick,M.H., Willard, H.F., Menlove, L.A.: Report of the Committee on Human Gene Mapping by Recombinant DNA Techniques. Human Gene Mapping (7) (1984): Seventh International Workshop on Human Gene Mapping. *Cytogenet. Cell Genet* 37(1-4):210-274, 1984.

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

4. Barker, D., Holm, T., White, R.: A locus on chromosome 11p with multiple restriction site polymorphisms. *Am. J. Hum. Genet* 36: 1159-1171,1984.

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

6. Rozen, R., Fox, J, Fenton, W.A., Horwich, A.L. and Rosenberg, L. E.: Gene deletion and restriction fragment length polymorphisms at the human ornithine transcarbamylase locus. *Nature* 313:815-817, 1985.

7. Buroker,N.E., Magenis, R.E., Wleiky K., Bruns, G., Litt, M.: Four restriction fragment length polymorphisms revealed by probes from a single cosmid map to human chromosome 12q. *Hum Genet* In press, 1986.

8.. Latt, S.A., Willard, H.F. and Gerald, P.S.: BrdU-33258 Hoechst analysis of DNA replication in human lymphocytes with supernumerary or structurally abnormal X chromosomes. *Chromosoma* 57: 135-153, 1976.

9. Brook, J.D., Shaw, D.J., Meredith, I. , Bruns, G.A.P. and Harper, P.S.: Localization of genetic markers and orientation of the linkage group on chromosome 19. *Hum. Genet* 68: 282-285,1984.

10. Bruns, G.A.P., Leary, A.C., Eisenman, R.E. Bazinet, C.W., Regina, V.M. and Gerald, P. S.: Expression of $ACON_S$ and GALT in man-rodent somatic cell hybrids. *Cytogenet. Cell Genet*22:172-176,1978.

11. Bruns, G.A.P., Mintz, B.J., Leary, A.C., Regina, V.M. and Gerald, P.S.: Human lysosomal genes -arylsulfatase A and β-galactosidase. *Biochem. Genet* 17:1031-1059, 1979.

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

13. Schweizer, D.: Simultaneous fluorescent staining of R-bands and specific heterochromatic regions (DA-DAPI bands) in human chromosomes. *Cytogenet. Cell Genet* 27:190-193, 1980.

14. Willard, H., Skolnick, M, Pearson, P., Mandel, J-L.: Report of the Committee on Human Gene Mapping by Recombinant DNA Techniques. Eighth International Workshop on Human Gene Mapping. *Cytogenet Cel, Genet* 40:360-489, 1985.

15. Szulman, A.E., Surti,U.: The syndromes of hydatidiform mole. I. Cytogenetic and morphologic correlations. *Am. J. Obstet. Gynecol* 131:665-671, 1978.

16. Westerveld, A. , Naylor,S.: Report of the Committee on the Genetic Constitution of Chromosomes 18, 19, 20 21 and 22. Human Gene Mapping (7) (1984): Seventh International Workshop on Human Gene Mapping. Cytogenet. Cell Genet 37(1-4):1-666., 1984.

17. Brook, J.D., Shaw, D.J., Meredith, A.L., Scott, J., Litt, M., Bufton, L., Harper,P.S.: A somatic cell hybrid mapping panel for chromosome 19: localisation of known genes and RFLPs and orientation of the linkage group. Human Gene Mapping (8) (1985): Eighth International Workshop on Human Gene Mapping. *Cytogenet. Cell Genet* 40:590, 1986.

18. Pericak-Vance, M.A., Bartlett, R.J., Ross, D.A., Fey, G., Bufton, L. Litt, M., Humphries, S., Williamson, R., Connealy, P.M., Roses, A.D.: Multipoint mapping of myotonic dystrophy with chromosome 19 serum, DNA and blood group polymorphisms. Human Gene Mapping (8) (1985): Eighth International Workshop on Human Gene Mapping. *Cytogenet. Cel. Genet* 40:721-722, 1986.

TABLE 1

Probe-Enzyme Pairs That Reveal RFLPs

Probe	Insert size (kb)	Enzyme	Fragment pattern	Allele frequencies (N)	Locus
1-82§	1.65	Taq I	4.3*,4.2* 2.5	0.38, 0.62 (110)	α
		Msp I	3.0*,2.9* 1.8, 0.67		α
		Hind III	6.2*,5.4*	0.15, 0.85 (110)	γ
		Bam HI	14.0*,11.2* 10.5*,1.0	0.10, 0.76, 0.14 (76)	δ
2-21	0.59	Msp I	2.6*, 2.5*	0.44, 0.56 (98)	β
		Bam HI	14.0*,11.2* 10.5*,6.4		δ

*Fragments observed to vary among unrelated individuals.

§Subclone 2-11, with a 0.63 kb insert, reveals the same polymorphisms as subclone 1-82, but does not reveal the constant fragments.

N, in parentheses, is the no. chromosomes studied.

TABLE 2

Chromosome Contents of Hybrid Cell Lines

The designations are: (+) presence or (-) absence of a human chromosome; (R), rearranged chromosome as determined by disruption of a syntenic group or by a cytogenetic abnormality; (±), chromosome present in less than 15% of cells and/or the isozyme or DNA probe characteristic of the chromosome weakly positive; (p), presence of a marker for the short arm and absence of a marker for the long arm; (q), presence of a marker for the long arm and absence of a marker for the short arm. For calculation of the discordancy fractions, hybrids with a rearranged chromosome or those where the chromosome was present in less than 15% of the cells were excluded from analysis. The column designated M indicates the presence or absence of the 4.2 and 2.8 kb Eco RI fragments in Southern transfers probed with subclone 1-82.

*Cell line G89E5 contains an X chromosome in all cells plus an unidentified human-hamster translocation in about 20% of the cells.

† On blots made from gels run longer for better resolution, the faint bands of size approximately 4.3 kb seen with cell lines G89E5, G95A4, G35AzA and G35AzF were seen to be distinctly slower than the human specific 4.2 kb band seen with G24B5.

	Chromosome Complement																									Scoring
Cell Line	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	Xq24-qter	Xpter-q24	Y	M
G35D5	+	+	+	-	-	+	+	-	-	+	-	-	+	+	+	+	+	+	+	+	-	+	+	-	-	+
G35F3	-	+	+	-	+	-	+	+	-	+	+	+	-	-	-	+	-	+	+	+	-	-	+	-	-	+
G35D2	±	-	-	-	+	+	-	-	+	-	+	+	-	-	+	+	-	+	R	+	-	+	R	-	-	-
G35E3	-	-	-	-	+	+	-	+	-	-	+	+	+	+	-	+	-	+	+	+	+	+	+	-	-	+
G35A2	-	-	+	+	-	+	-	-	-	-	+	-	-	+	-	-	-	-	+	+	+	+	+	-	-	+
G35C2	-	-	-	+	-	-	-	+	-	-	R	-	-	+	-	+	-	-	R	-	-	-	R	R	-	-
G35C4	-	-	-	-	-	+	-	-	R	-	-	+	-	+	-	-	-	±	+	+	+	-	+	-	-	+
G35B5	-	-	-	-	-	R	+	-	-	-	+	p	+	-	-	-	+	-	+	+	-	-	+	-	-	+
G35F5	±	-	+	+	-	+	-	-	+	-	+	-	+	+	-	+	+	+	+	±	+	+	+	+	-	+
G35D3	-	p	-	+	-	-	+	+	+	+	-	-	-	±	-	+	+	-	+	+	+	+	+	-	-	+
G35C5	-	-	p	-	+	-	+	+	+	+	-	p	-	-	+	+	+	-	+	+	+	+	+	-	-	+
G35A4	+	-	+	+	-	+	-	+	+	-	-	-	-	+	-	+	-	-	+	+	+	-	+	+	-	+
G35B4	+	-	-	-	+	-	-	-	-	+	+	+	-	-	-	-	-	+	+	-	-	+	+	+	-	+
G35E4	-	-	-	R	-	-	-	-	R	-	-	-	-	-	+	-	-	-	+	-	-	-	+	-	-	+
G175	±	-	-	+	-	+	-	+	-	+	-	-	+	+	+	-	-	-	+	-	-	+	+	-	-	+
G24A9	-	+	-	+	-	+	-	+	+	+	-	-	±	+	+	-	±	+	+	+	+	-	+	q	-	+
G24A4	±	p	-	-	+	+	+	-	±	-	±	-	+	+	+	-	-	-	+	-	+	-	+	q	-	+
G24B5	+	-	+	+	-	-	+	-	-	+	+	+	+	+	-	+	-	-	+	-	+	+	+	+	-	+
G89E5*	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	-	-[1]
G95A4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	±	-	-	+	+	-	-[1]
G35A5AzA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	ND	ND	-	-[1]
G35A5AzF	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	ND	ND	-	-[1]
RRP5-7	+	-	-	+	+	-	+	-	-	+	-	+	ND	+	-	-	+	+	-	+	-	-	+	+	R	-
Discordancy	10	11	9	10	13	7	10	10	9	9	9	11	8	7	11	10	10	10	0	7	6	8	3	13	16	
Fraction	19	23	23	22	23	22	23	23	20	23	21	23	21	22	23	23	21	21	21	21	23	23	19	18	22	

TABLE 3

Haplotype Frequencies

Haplotype	α	β	γ	δ	N*
I	2	1	2	2	13
II	1	2	2	2	9
III	2	2	1	2	4
IV	1	2	2	1	3
V	2	1	2	3	3
VI	1	1	2	1	1
VII	2	1	1	2	1
					34

*N is the no. chromosomes observed.

FIGURE 1

Southern transfers of restriction digests of a panel of DNAs from unrelated individuals probed with cosmid 1-13 after prehybridization of the cosmid with total human DNA. A, Taq I digests *(lanes a-e and g-j); lane f* has size markers. B, Msp I digests *(lanes a-e and g-j); lane f* has size markers. Reprinted from [3], with permission.

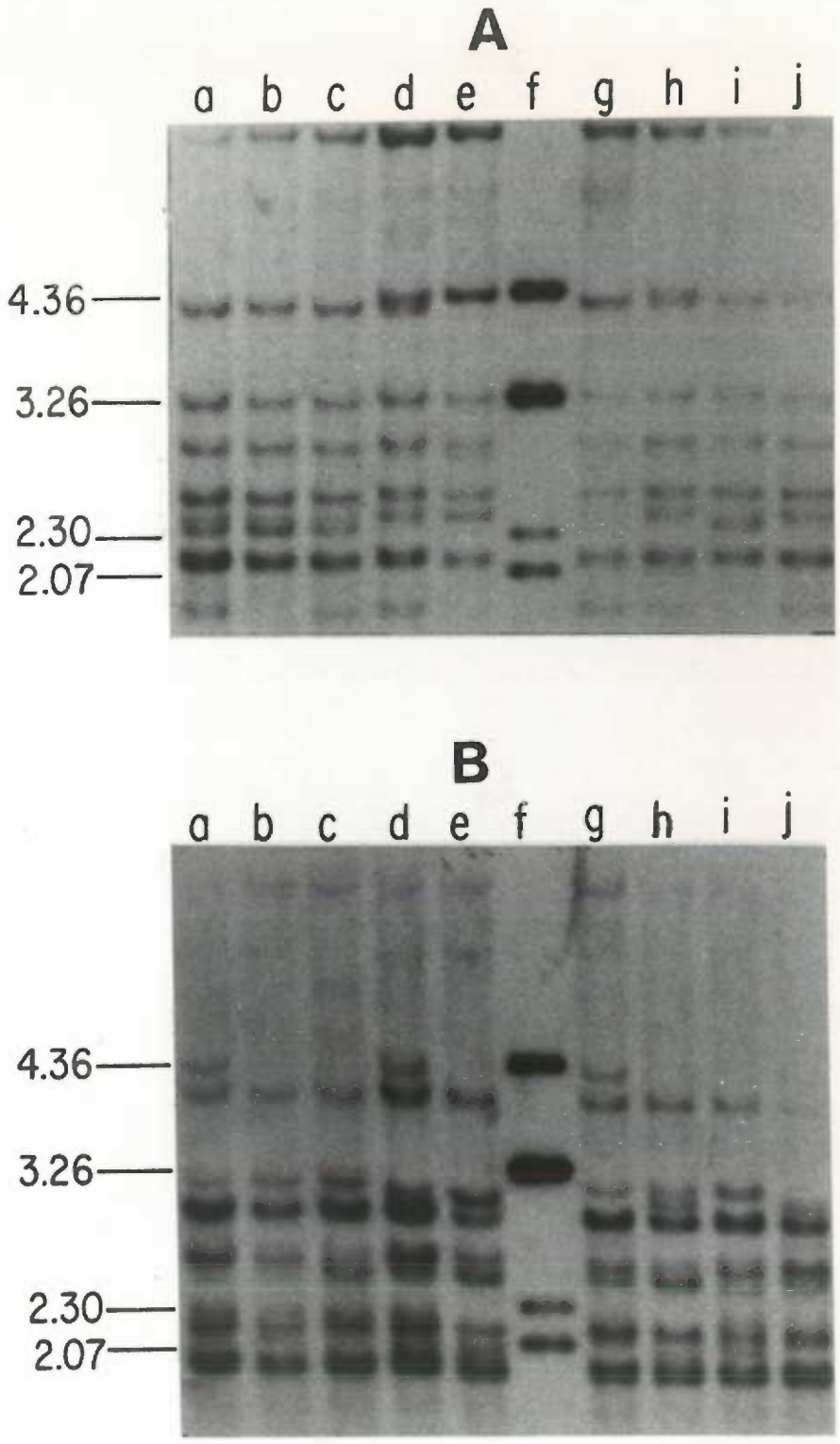
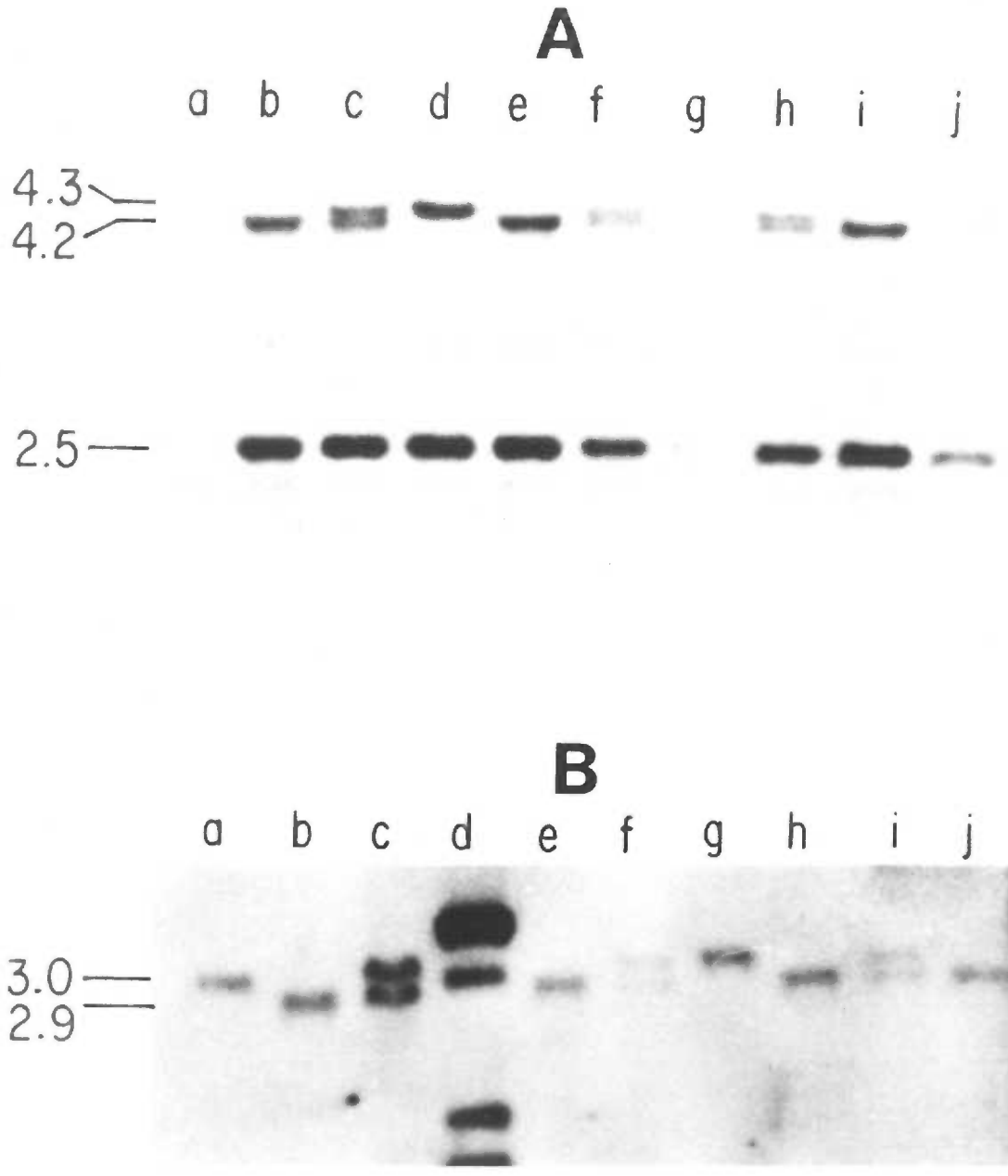
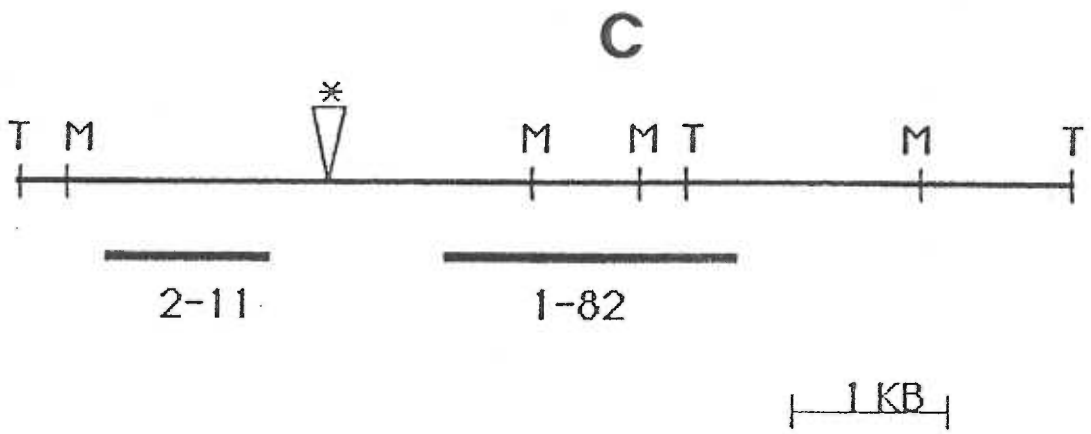

A
a b c d e f g h i j
4.36
3.26
2.30
2.07
B
a b c d e f g h i j
4.36
3.26
2.30
2.07

FIGURE2

RFLPs revealed by subclones of cosmid 1-13. A, Subclone 1-82 was used as a hybridization probe on Southern transfers of Taq I digested DNA from 9 unrelated individuals (*lanes b-j*). *Lane a* contains size standards. B, Subclone 2-11 was used as a hybridization probe on Southern transfers of Msp I digested DNA from 9 unrelated individuals (*lanes a-c, e-j*). *Lane d* contains size standards. C, Model for the α locus, showing a hypothetical map of Taq I (T) and Msp I (M) sites in the neighborhood of the 0.1 kb insertion (*). Possible locations of the inserts of subclones 1-82 and 2-11 are shown by *thick lines below the map*. D, Subclone 2-11 was used as a hybridization probe on Southern transfers of Hind III digested DNA from 11 unrelated individuals (*lanes a-k*). *Lane l* contains size standards. E, The insert from subclone 2-21 was used as a hybridization probe on Southern transfers of Msp I digested DNA from 6 unrelated individuals (*lanes b-g*). *Lane a* contains size standards.


A
a b c d e f g h i j
4.3
4.2
2.5
B
a b c d e f g h i j
3.0
2.9


C
T
M
*
M
M
T
M
T
2-11
1-82
1 KB


D
a b c d e f g h i j k l
6.2
5.4


E
a b c d e f g
2.6
2.5

FIGURE3

Subclones 1-82 (A) and 2-21(B) were used to probe Southern transfers of Bam HI-digested DNA from 3 unrelated individuals. The *leftmost lane* in each panel contains size standards.

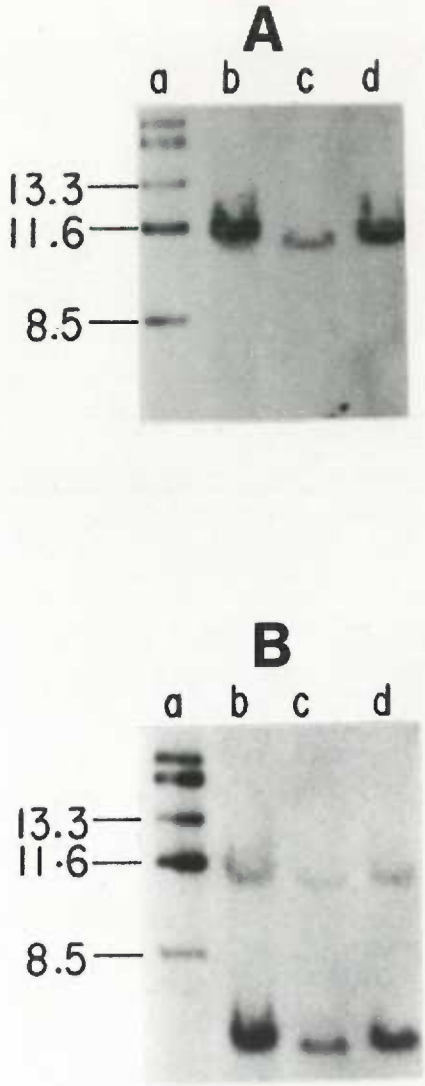

A
a
b
c
d
13.3
11.6
8.5
B
a
b
c
d
13.3
11.6
8.5

FIGURE 4

Southern blot of Eco RI digests of DNAs from parental and hybrid cell lines probed with subclone 1-82. *The lane farthest to the right* contains size standards.

FIGURE 5

Southern blot of Eco RI digests of DNAs from hybrid cell lines G175AoXi *(lane b),* G24B2AM (*lane c)* and GM89A99c7 *(lane d). Lane a* contains size markers.

4.2
2.8
E36
G35D5
G35F3
G35D2
G35E3
G35A2
G35C2
G35C4
G35B5
G35F5
G35D3
G35C5
G35A4
G35B4
G35E4
RAG
G175
G24A9
G24A4
G24B5
G89E5
G95A4
G35A5AZA
G35A5AZF
RRP5-7
a
b
c
d
4.2
2.8

FIGURE6

Ideogram of chromosome 19 showing chromosomal regions present in hybrid cell lines G175AoXi, G24B2AM, GM89A99c7 and indicating the shortest region of overlap (SRO) for localization of subclone 1-82.

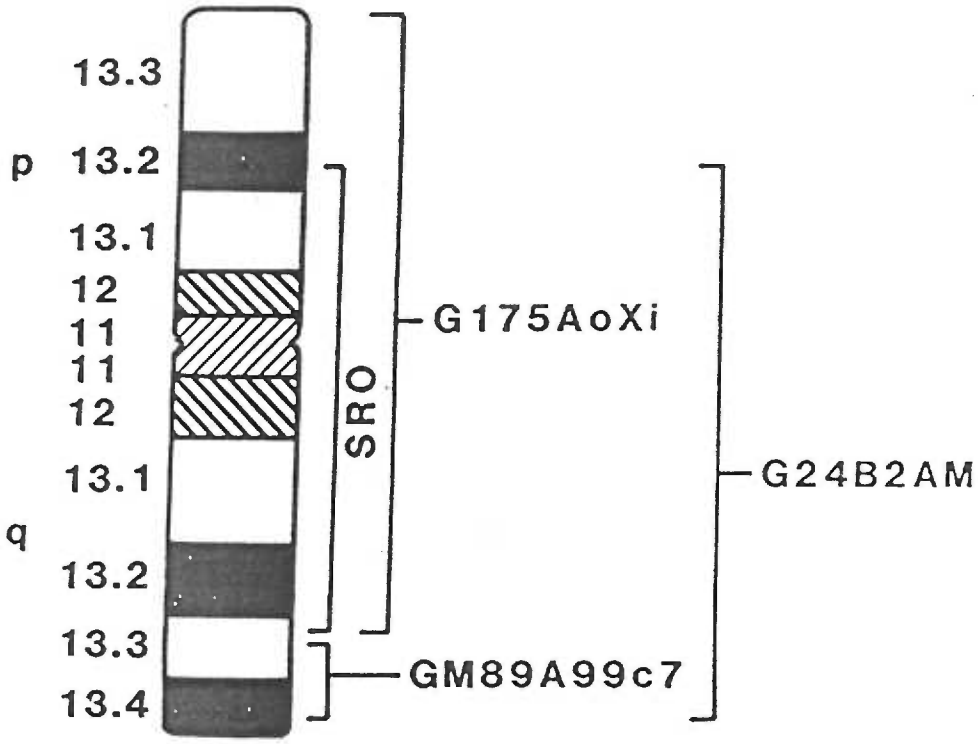
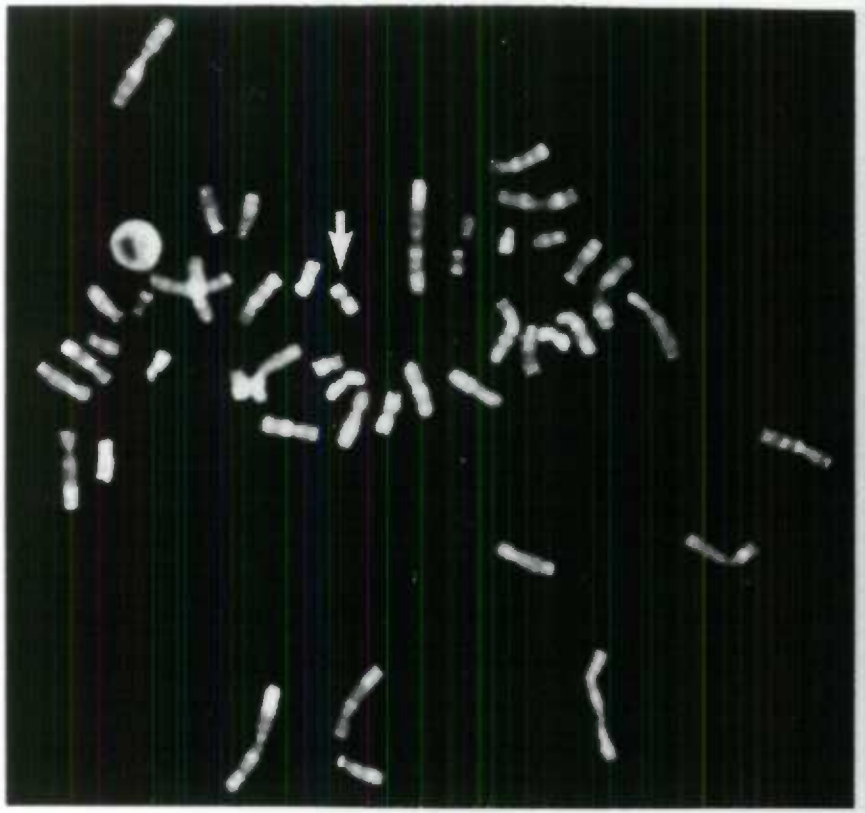

13.3
p 13.2
13.1
12
11
11
12
13.1
q
13.2
13.3
13.4
SRO
G175AoXi
G24B2AM
GM89A99c7

FIGURE7

In situ hybridization of probe 1-82. *Above,* representative metaphase which has been R-banded following hybridization. *Below,* the same metaphase, destained and restained with Wright's stain to show silver grains. *Arrows* indicate silver grain over 19p and the corresponding location in the R-banded metaphase.



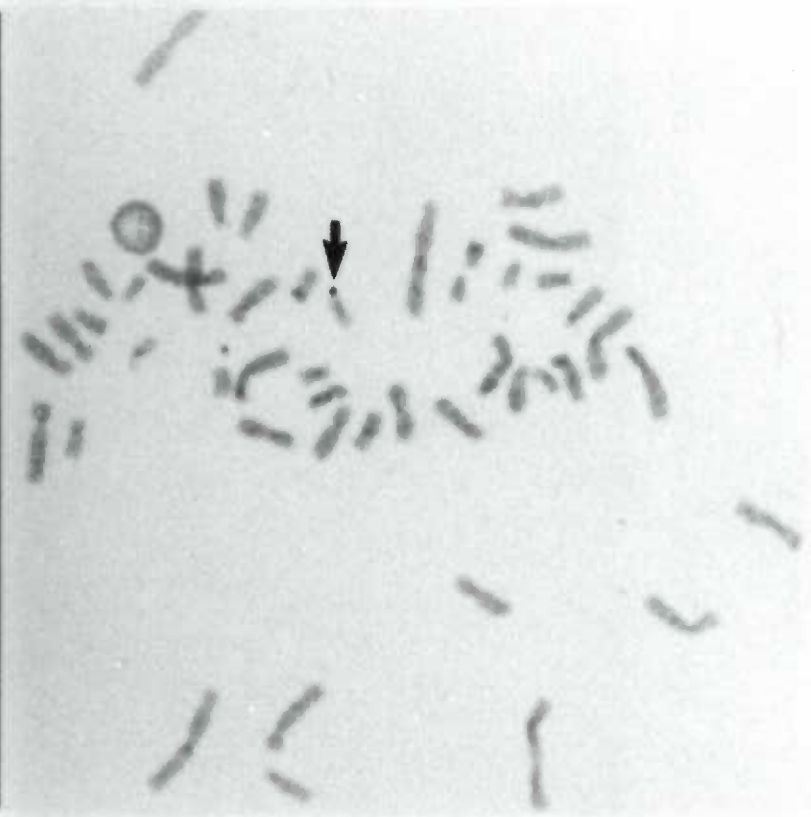
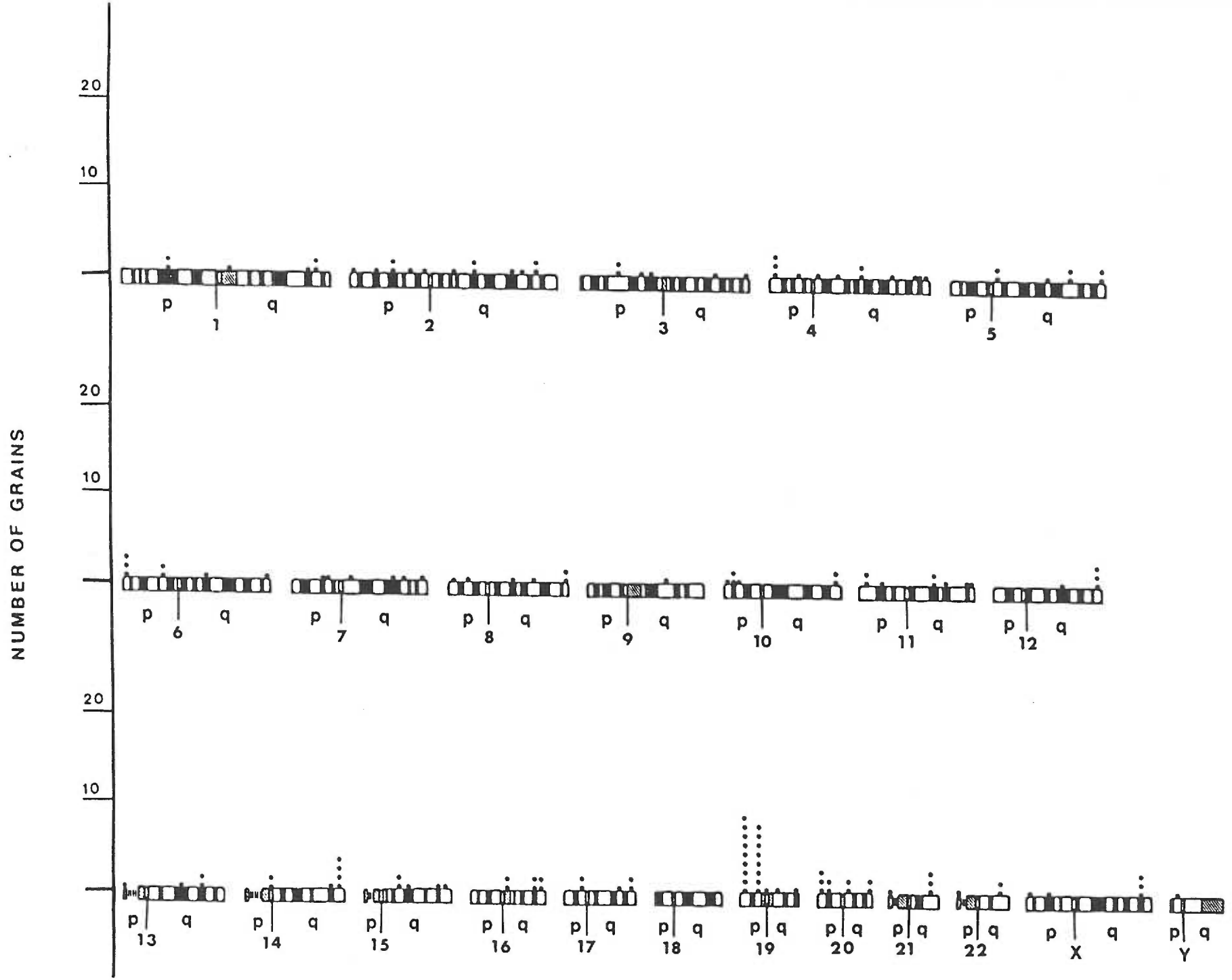


FIGURE8

Histogram of chromosomal distribution of silver grains from in situ hybridization of probe 1-82 to chromosomes of a normal male.


NUMBER OF GRAINS
20
10
20
10
20
10
p
q
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
X
Y

IV. SUMMARY AND CONCLUSIONS

The results of this research support the idea that cosmid libraries can be useful in the search for highly polymorphic RFLPs in the human genome.

From the 11q/16q library, two RFLPs were discovered and mapped to 16q. Probe p79-2-23 from cosmid CF33-79 reveals a highly polymorphic insertion/deletion locus with at least 11 alleles and 80% heterozygosity. Probe p391-2-42 from cosmid CF 32-391 reveals a one- or two-locus RFLP which requires further characterization.

A random cosmid, C1-13, from Ed Fristch's total human genomic library was studied and found to reveal a highly polymorphic compound locus on chromosome 19p, specifically localized to 19p13.2->19cen. This compound RFLP consists of at least 4 loci with a minimum 72% heterozygosity. Probes for this RFLP have been provided to other investigators working on a chromosome 19 map and on the degenerative muscle disease, myotonic dystrophy, also mapped to chromosome 19. They have also been provided to investigators using such RFLPs to follow engraftment after bone marrow transplantation.

This paper represents the first report of construction of a mouse-human somatic cell hybrid DNA cosmid library for RFLP searches and the preliminary screening of such a library with total mouse and total human DNA. The number of RFLPs found with this

partial 11q/16q library was much lower than anticipated, partly due to technical problems in colony screening and DNA isolation and possibly due to cosmid deletions.

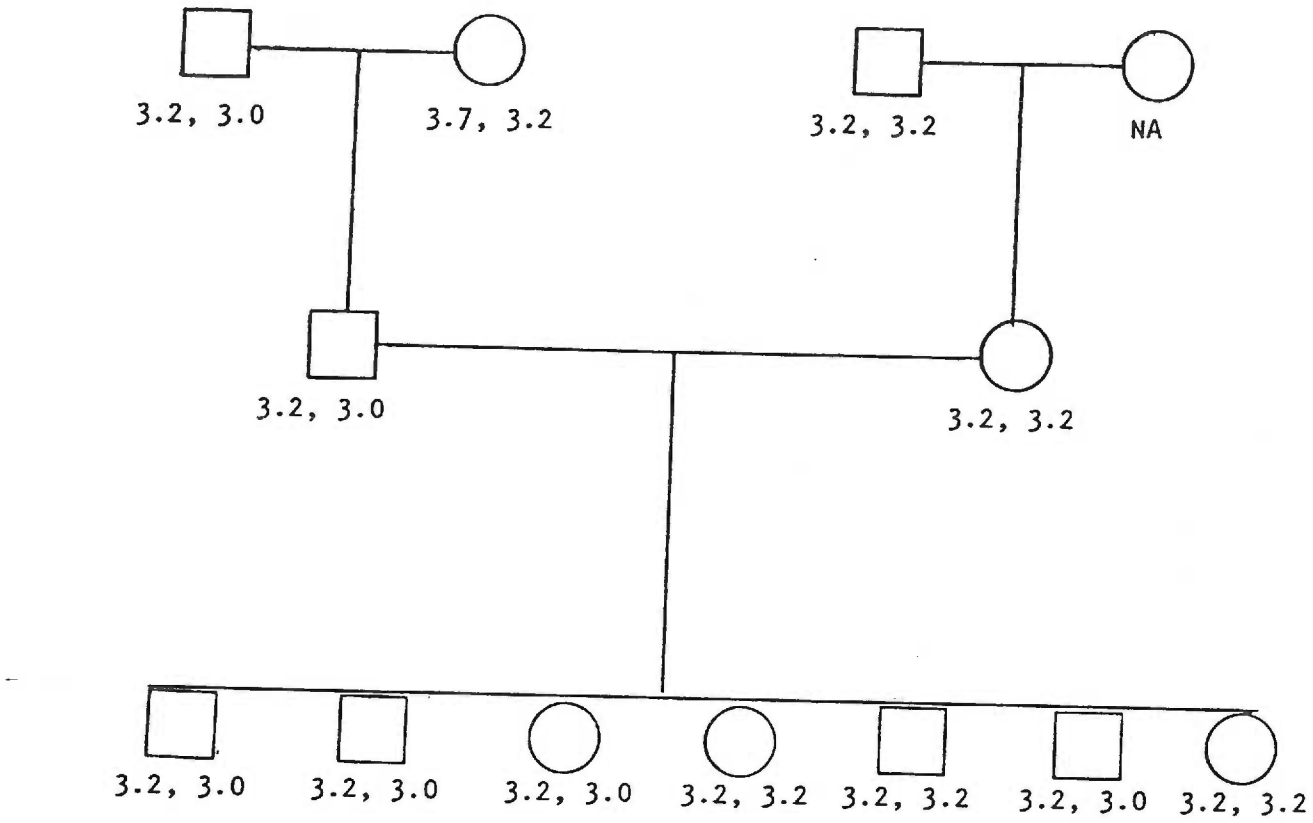
These problems need to be resolved if this method of RFLP searches is to be productive and time efficient. Many researchers have reported difficulties in the construction and stability of cosmid libraries, but with time and experience these problems are being resolved (Little & Cross, 1985; P. O'Connell, pers. comm., 1986; P. Little, pers. comm., 1986).

V. APPENDIX

APPENDIX A

CF33-79 RFLP Family Pedigrees

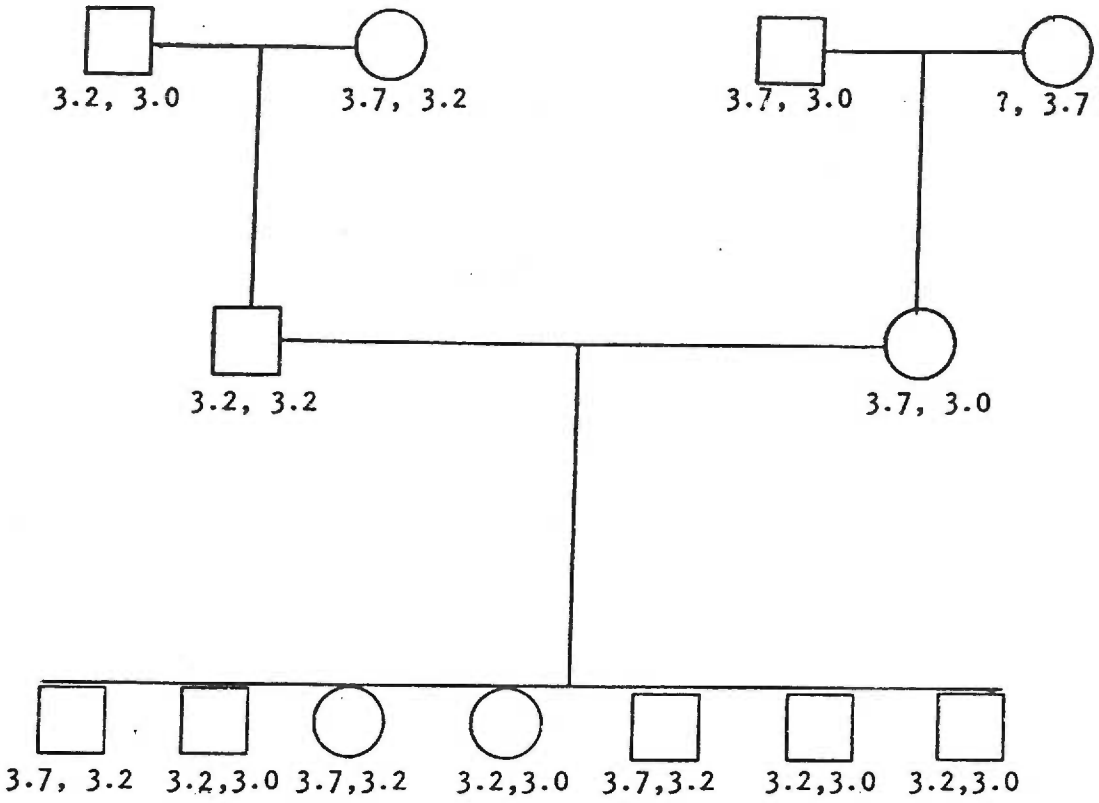
K 1329A



Taq I fragments

NA = not available

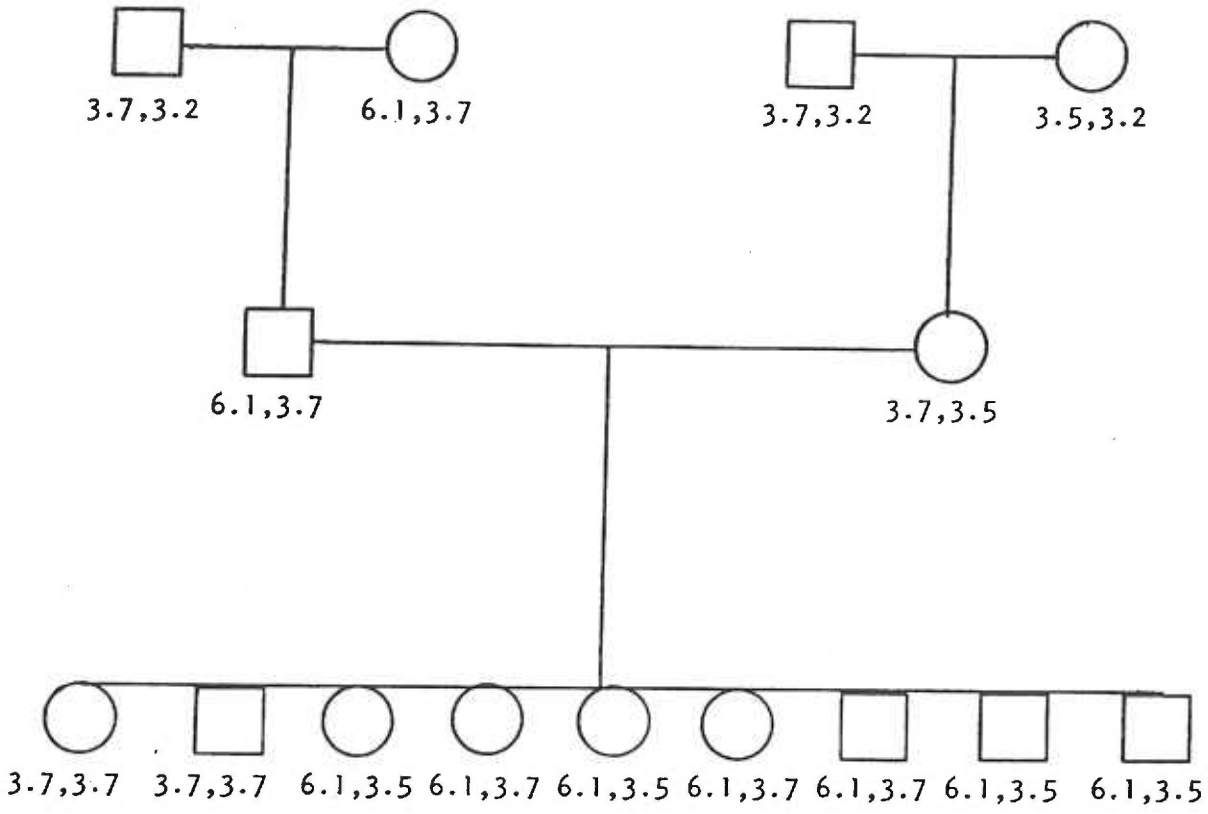
K 1329C



Taq I fragments

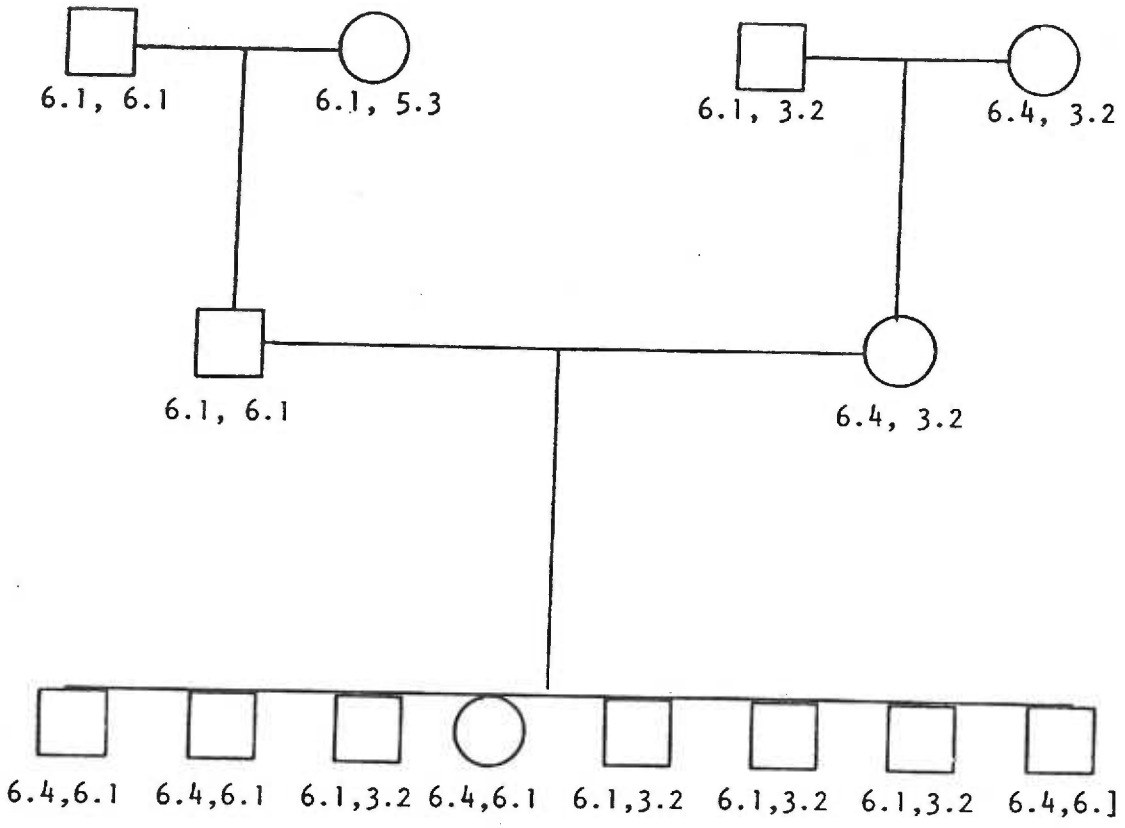
? = not readable on Southern blot

K 1331



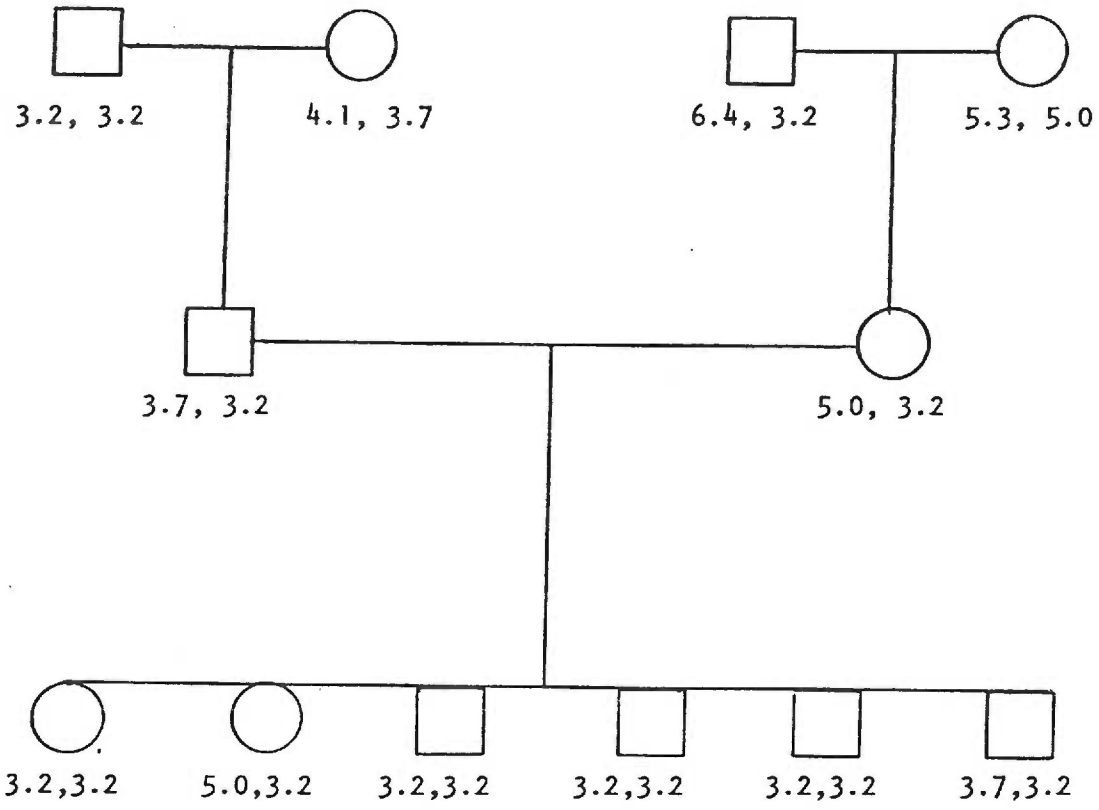
Taq I fragments

K 1333



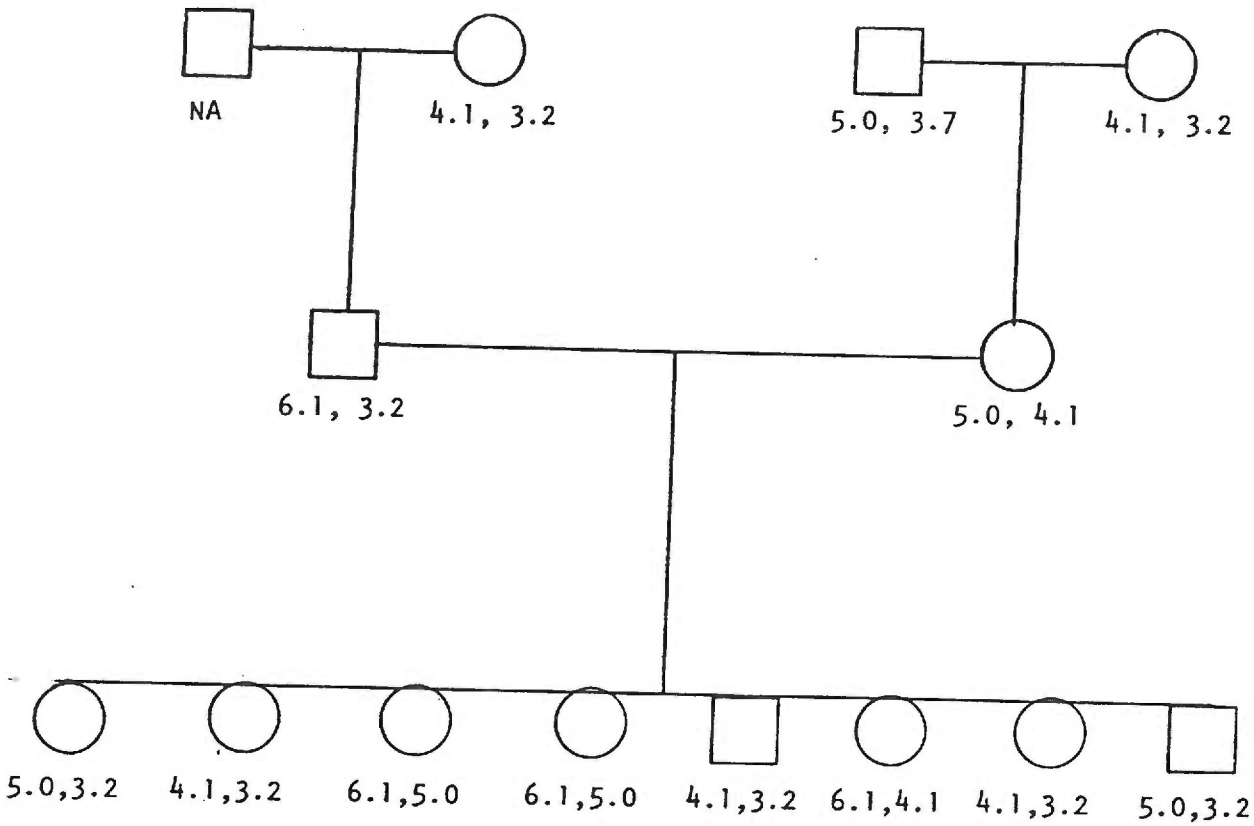
Taq I fragments

K 1340



Taq I fragments

K 1341



Taq I fragments

NA = not available

APPENDIX B - Solutions

acetate-MOPS: 0.1M sodium acetate, 0.5M 3-[N-Morpholino]propanesulfonic acid

alkaline SDS: 0.2N NaOH, 1% SDS

BCL (blood cell lysis) buffer: 0.32M sucrose, 10mM Tris Cl^-, pH7.5, 5mM $MgCl_2$, 1% Triton X-100

Denhardt's solution: 0.02% Ficoll 400, 0.02% polyvinylpyrollidone 360, 0.02% bovine serum albumin

dialysis solution A: 50mM Tris Cl^-, pH 8, 10mM EDTA, pH 8, 10mM NaCl

freezing medium: 2.5g yeast extract, 5g tryptone, 5g NaCl, 3.15g K_2HPO_4, 0.9g KH_2PO_4, 0.23g Na citrate, 0.45g $(NH_4)_2SO_4$, 2.2g glycerol, 0.045g $MgSO_4 \cdot 7H_2O$, H_2O to 500 ml

high salt solution: 3M potassium acetate, 1.8M formic acid

hybridization solution (with dextran sulfate): 5X SSC, 1X Dennhardt's, 20mM sodium phosphate, pH 6.5, 100 µg/ml sonicated denatured salmon DNA, 10% dextran sulfate, 50% formamide

LB (Luria-Bertania) agar: LB medium with 15g/L agar

LB (Luria-Bertania) medium: 10g Bacto-tryptone, 5g Bacto-yeast extract, 10g NaCl, H_2O to 1 liter, adjust pH to 7.5 with NaOH, autoclave one hour

ligation reaction mixture: 50mM Tris Cl^-, pH 7.4, 10mM $MgCl_2$, 1mM spermidine, 5mM dithiothreitol, 1mM ATP, 100µg/ml BSA

lysozyme solution: 50mM glucose, 25mM Tris Cl^-, pH 8, 10mM CTDA, pH 8

NL buffer: 75mM NaCl, 24mM EDTA, pH 8

OLB: 100:250:150 mixture of Solutions A, B, & C respectively. Solution A is1.25M Tris Cl^-, pH 8, 125mM $MgCl_2$, 250mM β-mercaptoethanol, 0.5mM dNTPs (minus dATP or dCTP depending on which radioactive dNTP was used); Solution B is 2M HEPES (N-2-hydroxyethylpiperazine-N'-2-ethanesulfonic acid); Solution C is random hexadeoxynucleotides (Pharmacia) dissolved at 90 O.D. units per ml in 1mM Tris Cl^- & 1mM EDTA, pH 7.5

prehybridization/hybridization solution (without dextran sulfate): 5X SSPE, 5X Denhardt's solution, 0.5 mg/ml sonicated denatured salmon DNA (Sigma), 50% formamide

prewash for colony filters: 1M NaCl, 50mM Tris Cl^-, pH 8, 1mM EDTA, pH 8, 0.1% SDS

RNAse: received from Sigma, dissolved at 10 mg/ml in H_2O, heated at 80^0C for 10 minutes to inactivate DNAse

SET: 0.15M NaCl, 30mM Tris Cl^-, pH 8, 1mM EDTA, pH 8

SM: 100mM NaCl, 50mM Tris Cl^-, pH 7.5, 0.01% gelatin, 8mM $MgSO_4$

SSC: 0.15M NaCl, 0.015M sodium citrate

SSPE: 0.75M NaCl, 50mM NaH_2PO_4, pH 7.4, 5mM EDTA, pH 7.4

TE: 10mM Tris Cl^-, 1mM EDTA, pH 8

TE^{-4}: 10mM Tris Cl^-, 0.1mM EDTA, pH 8

VI. REFERENCES

Antonarakis SE, Kazazian HH, Orkin SH: DNA polymorphism and molecular pathology of the human globin gene clusters. *Hum Genet* 69:1-14, 1985.

Bakker E, Goor N, Wrogemann K, Kunkel LM, Fenton WA, Majoor-Krakauer D, Jahoda MGJ, van Ommen GJB, Hofker MH, Mandel JL, Davies KE, Willard HF, Sandkuyl L, Essen AJV, Sachs ES, Pearson PL: Prenatal diagnosis and carrier detection of Duchenne muscular dystrophy with closely linked RFLPs. *Lancet* I:655-658, 1985.

Barker D, Schafer M, White R: Restriction sites containing CpG show a higher frequency of polymorphism in human DNA. *Cell* 36:131-138, 1984a.

Barker D, Holm T, White R: A locus on chromosome 11p with multiple restriction site polymorphisms. *Am J Hum Genet* 36:1159-1171, 1984.

Bell GI, Karam JH, Rutter WJ: Polymorphic DNA region adjacent to the 5' end of the human insulin gene. *Proc Natl Acad Sci USA* 78:5759-5763, 1981.

Benton WD, Davis RW: Screening λgt recombinant clones by hybridization to single plaques *in situ*. *Science* 196:180-182, 1977.

Birboim HC: A rapid alkaline extraction method for the isolation of plasmid DNA. *Methods in Enzymology* 100:243-249, 1983.

Bishop DT, Williamson JA, Skolnick MH: A model for restriction fragment length distributions. *Am J Hum Genet* 35:795-815, 1983.

Blazar BR, Orr HT, Arthur DC, Kersey JH, Filipovich AH: Restriction fragment length polymorphisms as markers of engraftment in allogenic marrow transplantation. *Blood* 66:1436-1444, 1985.

Botstein D, White RL, Skolnick M, Davis RW: Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet* 32:314-331, 1980.

Brook JD, Shaw DJ, Meredith I, Bruns GAP, Harper PS: Localization of genetic markers and orientation of the linkage group on chromosome 19. *Hum Genet* 68:282-285, 1984.

Bruns GAP, Leary AC, Eisenman RE, Bazinet CW, Regina VM, Gerald PS: Expression of ACONs and GALT in man-rodent somatic cell hybrids. *Cytogenet Cell Genet* 22:172-176, 1978.

Bruns GAP, Mintz BJ, Leary AC, Regina VM, Gerald PS: Human lysosomal genes - arylsulfatase A and β-galactosidase. *Biochen Genet* 17:1031-1059, 1979.

Bufton L, Bruns GAP, Magenis RE, Tomar D, Shaw D, Brook D, Litt M: Four restriction fragment length polymorphisms revealed by probes from a single cosmid map to chromosome 19. *Am J Hum Genet* 38, in press, 1986.

Buroker NE, Magenis RE, Weliky K, Bruns GAP, Litt M: Four restriction fragment length polymorphisms revealed by probes from a single cosmid map to human chromosome 12q. *Hum Genet* 72:86-94, 1986.

Busslinger M, deBoer E, Wright S, Grosveld FG, Flavell RA: The sequence GGC^{m}CGG is resistant to Msp I cleavage. *Nucleic Acids Res* 11:3559-3569, 1983.

Carbon J, Clarke L, Ilgen C, Ratzkin B: The construction and use of hybrid plasmid gene banks in *Escherichia coli*. In *Recombinant Molecules: Impact on Science and Society*, ed. by RF Beers and EG Bassett, New York, Raven Press, 1977 pp 335-378.

Cavenee WK, Dryja TP, Phillips RA, Benedict WF, Godbout R, Gallie BR, Murphree AL, Strong LC, White RL: Expression of recessive alleles by chromosomal mechanisms in retinoblastoma. *Nature* 305:779-784, 1983.

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

Collins J, Hohn B: Cosmids: a type of plasmid gene cloning vector that is packageable *in vitro* in bacteriophage λ heads. *Proc Nat. Acad Sci USA* 75:4242-4246, 1978.

Cooper DN, Smith BA, Cooke HJ, Niemann S, Schmidtke J: An estimate of unique DNA sequence heterozygosity in the human genome. *Hum Genet* 69:201-205, 1985.

Cox DW, Woo SLC, Mansfield T: DNA restriction fragments associated with α_1-antitrypsin indicate a single origin for deficiency allele PI Z. *Nature* 316:79-81, 1985.

Daniel A: The size of prometaphase chromosome segments. *Clin Genet* 28:216-224, 1985.

Dillela A, Woo S: Cosmid cloning of genomic DNA. *Focus* 7:1-5, 1985.

Dracopoli NC, Houghton AN, Old LJ: Loss of polymorphic restriction fragments in malignant melanoma: Implications for tumor heterogeneity. *Proc Natl Acad Sci USA* 82:1470-1474, 1985.

Drayna D, White R: The genetic linkage map of the human X chromosome. *Science* 230:753-758, 1985.

Feinberg AP, Vogelstein B: A technique for radiolabeling DNA restrtiction fragments to high specific activity. *Analyt Biochem.* 132:6-13, 1983.

Graham JB, Green PP, McGraw RA, Davis LM: Application of molecular genetics to prenatal diagnosis and carrier detection in the hemophilias: some limitations. *Blood* 66:759-764, 1985.

Grodzicker T, Williams J, Sharp P, Sambrook J: Physical mapping of temperature-sensitive mutations of adenoviruses. *Cold Spring Harbor Symp Quant Biol* 39:439-446, 1974.

Grosveld FG, Dahl H-HM, de Boer E, Flavell RA: Isolation of β-globin related genes from a human cosmid library. *Gene* 13:227-237, 1981.

Grunstein M, Hogness DS: Colony hybridization: a method for the isolation of cloned DNAs that contain a specific gene. *Proc Nat. Acad Sci USA* 72:3961-3965, 1975.

Gumucio DL, Wiebauer K, Dranginis A, Samuelson LC, Treisman LO, Caldwell RM, Antonucci TK, Meisler MH: Evolution of the amylase multigene family. *J Biol Chem* 260:13483-13489, 1985.

Gusella JF, Keys C, Varsanyi-Breiner A, Kao F-T, Jones C, Puck TT, Housman D: Isolation and localization of DNA segments from specific human chromosomes. *Proc Natl Acad Sci USA* 77:2829-2833, 1980.

Gusella JF, Wexler NS, Conneally PM, Naylor SL, Anderson MA, Tanzi RE, Watkins PC, Ottina K, Wallace MR, Sakaguchi AY, Young AB, Shoulson I, Bonilla E, Martin JB: A polymorphic DNA marker genetically linked to Huntington's disease. *Nature* 306:234-238, 1983.

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

Harper ME, Saunders GF: Localization of single-copy DNA sequences on G-banded chromosomes by in situ hybridization. *Chromosoma* 83:431-439, 1981.

Harris H, Hopkinson DA: Average heterozygosity per locus in man: an estimate based on the incidence of enzyme polymorphisms. *Ann Hum. Genet* 36:9-20, 1972.

Harris H: *The Principles of Human Biochemical Genetics*, 3rd ed. Amsterdam, Elsevier/North-Holland Biomedical Press, 1980.

Hoopes BC, McClure WR: Studies on the selectivity of DNA precipitation by spermine. *Nucleic Acids Res* 9:5493-5504, 1981.

Hutchison CA, Newbold JE, Potter SS, Edgell MH: Maternal inheritance of mammalian mitochondrial DNA. *Nature* 251:536-538, 1974.

Ish-Horowicz D, Burke JT: Rapid and efficient cosmid cloning. *Nucleic Acids Res* 9:2989-2998, 1981.

Jeffreys AJ: DNA sequence variants in the G_γ-, A_γ-, δ- and β-globin genes of man. *Cell* 18:1-10, 1979.

Jeffreys AJ, Wilson V, Thein SL: Hypervariable 'minisatellite' regions in human DNA. *Nature* 314:67-73, 1985a.

Jeffreys AJ, Wilson V, Thein SL: Individual-specific 'fingerprints' of human DNA. *Nature* 316:76-79, 1985b.

Jeffreys AJ, Brookfield JFY, Semeonoff R: Positive identification of an immigration test case using human DNA fingerprints. *Nature* 317:818-819, 1985c.

Kan YW, Dozy AM: Antenatal diagnosis of sickle-cell anaemia by D.N.A. analysis of amniotic fluid cells. *Lancet* 2:910-911, 1978.

King PV, Blakesley RW: Optimizing DNA ligations for transformation. *Focus* 8:1-3, 1986.

Knowlton RG, Cohen-Haguenauer O, Cong NV, Frezal J, Brown VA, Barker D, Braman JC, Schumm JW, Tsui L-C, Buchwald M, Donis-Keller H: A polymorphic DNA marker linked to cystic fibrosis is located on chromosome 7. *Nature* 318:380-382, 1985a.

Knowlton R, Schumm J, Braman J, Barker D, Donis-Keller H: Relative efficiencies of restriction endonucleases in detecting RFLPs. Eighth International Workshop on Human Gene Mapping. *Cytogenet Cell Genet* 40:670, 1985b.

Kolata G: Closing in on the muscular dystrophy gene. *Science* 230:307-308, 1985.

Koufos A, Hansen MF, Lampkin BC, Workman ML, Copeland NG, Jenkins NA, Cavenee WK: Loss of alleles at loci on human chromosome 11 during genesis of Wilms' tumour. *Nature* 309:170-172, 1984.

Kunkel LM, Monaco AP, Middlesworth W, Ochs HD, Latt SA: Specific cloning of DNA fragments absent from the DNA of a male patient with an X chromosome deletion. *Proc Natl Acad Sci USA* 82:4778-4782, 1985.

Lange K, Boehnke M: How many polymorphic genes will it take to span the human genome. *Am J Hum Genet* 38:842-845, 1982.

Lathrop GM, Lalouel JM, Julier C, Ott J: Strategies for multilocus linkage analysis in humans. *Proc Natl Acad Sci USA* 81:3443-3446, 1984.

Lathrop GM, Lalouel JM, White RL: Construction of human linkage maps: Liklihood calculation for multilocus linkage analysis. *Genet Epid* 3:39-52, 1986.

Latt SA, Willard HF, Gerald PS: BrdU-33258 Hoechst analysis of DNA replication in human lymphocytes with supernumerary or structurally abnormal X chromosomes. *Chromosoma* 57:135-153, 1976.

Lawn RM, Efstatiadis A, O'Connell C, Maniatis T: The nucleotide sequence of the human β-globin gene. *Cell* 21:647-651, 1980.

Lewin B: *Genes*, 2nd ed. New York, John Wiley & Sons, 1985.

Lindenmaier W., Dittmar K, Hauser H, Necker A, Sebald W: Isolation of a functional human interleukin 2 gene from a cosmid library by recombination *in vivo*. *Gene* 39:33-39, 1985.

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

Litt M, Bruns GA, Sheehy R, Magenis RE: A highly polymorphic locus in human DNA revealed by probes from cosmid 1-5 maps to chromosome 2q35->37. *Am J Hum Genet* 38:288-296, 1986.

Little PFR, Cross SH: A cosmid vector that facilitates restriction enzyme mapping. *Proc Natl Acad Sci USA* 82:3159-3163, 1985.

Magenis RE, Hecht F, Lovrien EW: Heritable fragile site on chromosome 16: Probable localization of haptoglobin locus in man. *Science* 170:85-87, 1970.

Maniatis T, Hardison RC, Lacy E, Lauer J, O'Connell C, Quon D: The isolation of structural genes from libraries of eucaryotic DNA. *Cel.* 15:687-701, 1978.

Maniatis T, Fritsch EF, Sambrook J: *Molecular Cloning. A Laboratory Manual.* New York, Cold Spring Harbor Laboratory,1982.

Marx JL: Putting the human genome on the map. *Science* 29:150-151, 1985.

Monaco AP, Bertelson CJ, Middlesworth W, Colletti C-A, Aldridge J, Fischbeck KH, Bartlett R, Pericak-Vance MA, Roses AD, Kunkel LM: Detection of deletions spanning the Duchenne muscular dystrophy locus using a tightly linked DNA segment. *Nature* 316:842-845, 1985.

Nathan D, Smith HO: Restriction endonucleases in the analysis and restructuring of DNA molecules. *Ann Rev Biochem* 44:273-293, 1975.

Naylor S, Lalouel JM, Shaw DJ: Report of the committee on the genetic constitution of chromosomes 17, 18, and 19. Eighth International Workshop of Human Gene Mapping. *Cytogenet Cell Gene* 40:242-267, 1985.

Oberle I, Camerino G, Heilig R, Grunebaum L, Cazenave J-P, Crapanzano C, Mannucci PM, Mandel J-L: Genetic screening for hemphilia A(classic hemophilia) with a polymorphic DNA probe. *N Engl J Med* 312:682-686, 1985.

Old RW, Primrose SB: *Principles of Gene Manipulation* 3rd ed. Oxford, Blackwell Scientific Publications, 1985.

Ott J: *Analysis of Human Genetic Linkage* Baltimore, The Johns Hopkins University Press, 1985.

Reeders ST, Breuning MH, Davies KE, Nicholls RD, Jarman AP, Higgs DR, Pearson PL, Weatherall DJ: A highly polymorphic DNA marker linked to adult polycystic kidney disease on chromosome 16. *Nature* 317:542-544, 1985.

Renwick JH: The mapping of the human chromosomes. *Ann Rev Genet* 5:81-120, 1971.

Rosbash M, Campo MS, Gummerson KS: Conservation of cytoplasmic poly (A)-containing RNA in mouse and rat. *Nature* 258:682-686, 1975.

Savatier P, Trabuchet G, Faure C, Chebloune Y, Gouy M, Verdier G, Negon VM: Evolution of the primate β-globin gene region. High rate of variation in CpG dinucleotides and in short repeated sequences between man and chimpanzee. *J Mol Biol* 182:21-29, 1985.

Schleif RF, Wensink PC: *Practical Methods in Molecular Biology.* New York, Springer-Verlag, 1981.

Schweizer D: Simultaneous flourescent staining of R-bands and specific heterochromatic regions (DA-DAPI bands) in human chromosomes. *Cytogenet Cell Genet* 27:190-193, 1980.

Sharp PA, Sugden B, Sambrook J: Detection of two restriction endonuclease activities in *Haemophilus parainfluenzae* using analytical agarose-ethidium bromide electrophoresis. *Biochemistry* 12:3055-3063, 1973.

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

Swartz MN, Trautner TA, Kornberg A: Enzymatic synthesis of deoxyribonucleic acids. *J Biol Chem* 237:864-875, 1962.

Urlaub G, Carothers AM, Chasin LA: Efficient cloning of single-copy genes using specialized cosmid vectors: Isolation of mutant dihydrofolate reductase genes. *Proc Natl Acad Sci USA* 82:1189-1193, 1985.

Vogelstein B, Gillespie D: Preparative and analytical purification of DNA from agarose. *Proc Natl Acad Sci USA* 76:615-619, 1979.

Vogelstein B, Fearon ER, Hamilton SR, Feinberg AP: Use of restriction fragment length polymorphisms to determine the clonal origin of human tumors. *Science* 227:642-644, 1985.

Wainwright BJ, Scambler PJ, Schmidtke J, Watson EA, Law H-Y, Farrall M, Cooke HJ, Eiberg H, Williamson R: Localization of cystic fibrosis locus to human chromosome 7cen-q22. *Nature* 318:384-385, 1985.

Weinberg RA: The molecules of life. *Sci Amer* 253:48-57, 1985.

Wensink PC, Finnegan DJ, Donelson JE, Hogness DS: A system for mapping DNA sequences in the chromosomes of *Drosophila melanogaster*. *Cell* 3:315-325, 1974.

White RL: DNA in medicine. Human genetics. *Lancet* II:1257-1262, 1984.

White R, Leppert M, Bishop T, Barker D, Berkowitz J, Brown C, Callahan P, Holm T, Jeromenski L: Construction of linkage maps with DNA markers for human chromosomes. *Nature* 313:101-105, 1985a.

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

Willard HF, Skolnick MH, Pearson PL, Mandel J-L: Report of the committee on human gene mapping by recombinanat DNA techniques. Eighth International Workshop of Human Gene Mapping. *Cytogenet Cel. Genet* 40:360-490, 1985.

Woo SLC, Lilsky AS, Guttler F, Chandra T, Robson KJH: Cloned human phenylalanine hydroxylase gene allows prenatal diagnosis and carrier detection of classical phenylketonuria. *Nature* 306:151-155, 1983.

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

Yang TP, Patel PI, Chinault AC, Stout JT, Jackson LG, Hildebrand BM, Caskey CT: Molecular evidence for new mutation at the *hprt* locus in Lesch-Nyhan patients. *Nature* 310:412-414, 1984.