THE CLONING AND CHARACTERIZATION OF DNA PROBES WHICH DETECT RESTRICTION FRAGMENT LENGTH POLYMORPHISMS ON HUMAN CHROMOSOMES 11q AND 16q

by

Cheryl L Maslen

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

May 1987

APPROVED:

[redacted]

(Professor in Charge of Thesis)

[redacted]

(Chairman, Graduate Council)

In memory of my grandmother, Edna Wardius, whose expressions of love and pride still echo through my mind.

ACKNOWLEDGMENTS

I will always be grateful to my thesis advisor, Dr. Michael Litt, who was enthusiastic and available; providing guidance when needed, but allowing me the independence and freedom so necessary to learning.

I would also like to thank those members of our research group, past and present, that I have been fortunate enough to work with: Dr. Linda Bufton, Ruth Litt, Jill Kellogg, Sophia Kondoleon, and Elaine Ostrander.

My research would not have been complete without the fine work of Bob Sheehy. Best of luck in graduate school.

I would like to thank John Adelman for not teaching me *everything* he knows about the bacteriophage lambda, but for patiently providing me with the information I've needed.

I'd also like to thank Chris Bond for her valuable assistance in many technical matters.

I have much appreciation for the computer expertise of Christopher Dubay, not to mention his time and patience in dealing with reams of data.

There have been many instances over the past several years that I have relied on the support and advice of Drs. Robert Koler, Richard Jones, and Ellen Magenis. Thank you for your time, encouragement and guidance.

My appreciation to the Coca-Cola company for leaving the caffeine in Diet Coke.

To my friends and colleagues Elaine Ostrander and Sophia Kondoleon:

Elaine, your friendship is one of the most valuable things I leave graduate school with. Thank you for sharing so much with me.

Soph, knowing you is knowing kindness and thoughtfulness. There is no way I can thank you enough for all you've done.

To my family:
I realize how unique the love and support is that I receive from all of you; especially you, Mom and Dad. You've made this possible in so many ways. Thank you.

And finally to David:
How can I thank you for the love, the sacrifices, the never-faltering support and overwhelming confidence you've shown in me. I look forward to our years ahead with greatest anticipation.

CONTENTS

TABLES

Page

SECTION II.

SECTION III.

SECTION IV.

FIGURES

Page

SECTION I.

SECTION II.

SECTION III.

ABBREVIATIONS

A	adenine
APRT	adenine phosphoribosyl transferase
bp	base pairs
C	cytosine
CEPH	*Centre d'Etude du Polymorphisme Humain* (Human Polymorphism Study Center)
cM	centiMorgans
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphates
dpm	disintegrations per minute
G	guanine
HGM	Human Gene Mapping Workshop
kb	kilobase pairs
M	molar
mg	milligram
ml	milliliter
mM	millimolar
mw	molecular weight
µg	microgram
µl	microliter

ng	nanogram
p	plasmid
PIC	polymorphism information content
RFLP	restriction fragment length polymorphism
RNA	ribonucleic acid
SM	storage media (Maniatis *et al*, 1982)
ss	size standard
T	thymine
Φ	phage
2xYT	enriched growth media (Silhavy *et al*, 1984)

ABSTRACT

Restriction fragment length polymorphisms (RFLPs) provide access to the normal diversity present in the human genome. The detection of heritable differences as reflected by RFLPs provides a basis for linkage analysis, thus allowing identification of genetic traits at the DNA level. Consequently, construction of genetic maps using RFLPs as marker loci facilitates the placement of genes and inherited traits by linkage analysis.

Human chromosomal regions 11q and 16q are particularly devoid of known genetic markers. This thesis strategically targets those areas for the purpose of regionalized mapping of RFLP loci as genetic markers. A somatic cell hybrid retaining a human 11q;16q translocation as the sole human DNA component was used to target these regions. A chromosome specific phage library was constructed from this preselected DNA as a source for the isolation of recombinant DNA probes which detect polymorphic loci. These loci will serve, in part, as the basis for linkage maps of these chromosomal regions.

As a result, a DNA probe which detects a novel cluster of at least 15 restriction site polymorphisms, deriving from chromosomal region 16q22-24, was cloned and characterized as to genetic and physical parameters.

There is no evidence that an insertion/deletion mechanism is responsible for this high degree of polymorphism. Furthermore, restriction analyses indicate that mutations at CpG dinucleotides did not play a predominant role in the formation of this unusual cluster of restriction site polymorphisms. Consideration of just three of the loci provides a highly polymorphic compound marker for that genomic region, with a population heterozygosity of 86%. Further characterization of the cluster will yield a marker with a heterozygosity approaching 100%. In addition, the novel genomic region identified by this probe will provide an opportunity to examine any relationship between the linkage disequilibium seen with some of the closely linked loci and the physical distance separating those loci.

Six probes which reveal RFLPs from chromosomal region 11q were also isolated and characterized, including three which detect closely linked compound loci. Physical mapping of the loci by use of a somatic cell hybrid panel shows that the probes span the region from 11q22 to 11qter. One probe location was confirmed by *in situ* hybridization. Linkage analyses for all pairwise combinations provides genetic evidence of linkage consistent with the physical map. In addition, there is genetic evidence of close linkage between two of the compound loci mapping to 11q22, which could result in

a population heterozygosity of at least 90% for the newly created compound locus. The isolation and localization of these probes provides a substantial basis for the construction of a genetic linkage map for chromosome region 11q.

I. INTRODUCTION

A. *Polymorphism in the Human Genome*

The human genome is a complex system with significant variation in the deoxyribonucleic acid (DNA) nucleotide sequences. Some of the variability in DNA sequence results in the virtually unlimited range of phenotypic differences observed in all people. An estimated 28% of recognized coding regions in the human genome have multiple forms which result in allelic expression of the gene, while silent or phenotypically unrecognized deviations exist in many others (Harris and Hopkinson, 1972). Dissimilarities of sequence are also known to occur in the noncoding regions of DNA, which cover as much as 80-90% of the genome (Lewin, 1983). In fact, sequence variation is hypothesized to be more frequent in these regions due to a lack of selective constraints in the absence of functionality (Lewin, 1983). Approximately one in every 100 to 200 base pairs (bp) of noncoding DNA is variable (Francomano and Kazazian, 1986). The identification of these regions occurs at the molecular rather than phenotypic level.

DNA sequence variations can be further divided into two categories which depend on the frequency of the variant in a given population. Heritable variations for which at least one to ten percent of chromosomes carry the minor allele have been traditionally termed polymorphisms (Lewontin, 1967). The definition is arbitrary and will be

taken as 90% for the most frequent allele for purposes of this thesis. Variations which occur less frequently in a given population are known as rare alleles. Originally these definitions applied to protein variants, but recently they have been extended to include genetic differences that can only be detected at the DNA level.

Prior to molecular detection of DNA variants, discrete heritable differences were detected as protein polymorphisms and rare alleles. In general, with the exception of single gene defects, phenotypic differences arise from complex genetic or genetic and environmental interactions, making it impossible to infer variation at a single locus. As a result, protein polymorphisms, detected as electrophoretic or antigenic variants, were the principal tool of geneticists interested in studying the human genome. Protein polymorphisms and cytogenetically visible chromosome heteromorphisms provided the first genetic markers which allowed chromosome assignment and genetic map location for single gene traits through linkage analysis.

The frequency of recombination between two observable traits provides information about the genetic distance between the loci on a given chromosome. Genetic distance is expressed in centiMorgans (cM) and loci which recombine once in every 100 meiotic opportunities are considered to be one cM apart (Morton, 1962).

Demonstration of genetic linkage between the Duffy blood group locus and a secondary constriction of chromosome

1, 1qh, provided the first autosomal assignment of a gene (Donahue *et al.*, 1968). This and subsequent assignments established a chromosomal map of genetic markers useful to ensuing linkage studies.

Linkage analysis has proven to be particularly applicable to inherited disorders resulting from dominant or codominant single gene defects. Appraisal of linkage is straightforward as coinheritance of a marker and phenotype can be followed in family studies. Recessive traits present a more difficult situation since the various alleles may not be detectable in the heterozygous carrier. Knowledge or inference of genotypes is essential to family studies for the assessment of cosegregation of phenotypic traits with specific marker alleles. Genetic heterogeneity, or the consequence of a phenotypic trait which can actually arise through modulation at more than one genetic locus, also complicates linkage analyses. Detection of linkage between a phenotype and marker in one family, but disagreement in another, may be an indicator of heterogeneity rather than a true lack of linkage.

A gene can be mapped through linkage studies using established markers if it is fortuitously near a known locus. Alternatively, failure to detect linkage between a trait and a marker locus is a powerful means of eliminating a specific genetic region or a candidate gene from consideration. However, the numbers and distribution of protein polymorphisms are severely limited as they derive

only from DNA coding sequences. These limitations leave large portions of the human genome devoid of marker loci, and consequently unmappable with this technology.

B. *Introduction to Restriction Fragment Length Polymorphisms*

All genetic variation is ultimately due to primary sequence variation at the DNA level. The utility of polymorphisms occurs only when the differences between the two copies of a given locus can be easily detected. The normal variation that exists in the human genome must be readily accessible before it can provide a basis for a genetic linkage map.

Recombinant DNA technology has provided the tools through which DNA from any source can be easily manipulated in a variety of ways (Maniatis *et al.*, 1982). Type II restriction endonucleases, commonly called restriction enzymes, can be used to cleave DNA at sequence specific sites resulting in DNA fragments of defined lengths. These so called restriction fragments are easily detected when hybridized with cloned probes homologous to those sequences (Southern, 1975).

If DNA is cleaved at a restriction enzyme recognition site in which polymorphic sequence variation has occurred, fragments of varying length result due to the presence or absence of the cleavage site (Fig.1a). Similarly, if DNA is cut with a restriction enzyme in a region where the sites of cleavage are separated by varying distances due to the

polymorphic insertion or deletion of DNA sequences, variation in resulting fragment lengths also occurs (Fig.1b). DNA fragments which derive from these events have been termed restriction fragment length polymorphisms (RFLPs). RFLPs are frequently categorized by the type of molecular mechanism from which they were generated. Both restriction site and insertion/deletion polymorphisms are stably inherited as codominant Mendelian alleles.

C. *Levels of DNA Polymorphism*:

1. Frequency in the Human Genome

An estimate of the number of RFLP loci in the human genome has been made based on the reassociation kinetics of DNA from a single study. Botstein *et al.* (1980), estimated that restriction enzymes with four bp recognition sequences would generate RFLPs at a rate of 0.0034 to 0.123 per restriction fragment of an average size of 256 bp. Restriction enzymes that have six bp recognition sequences should produce variable length fragments 0.005 to 0.177 times for each restriction fragment of an average length of 4096 bp. This method neglects the insertion/deletion polymorphisms and is therefore a conservative estimate. Empiric data estimates the frequency of DNA polymorphisms detectable as RFLPs to be approximately 0.7 to 0.8%, which is consistent with the values calculated by Botstein *et al.* (Jeffreys, 1979; Cooper *et al.*, 1985).

Some evidence indicates that point mutations occur preferentially at CpG dinucleotides because of the increased mutation rate at 5-methylcytosines (Barker *et al.*, 1984; Youssoufian *et al.*, 1986). Spontaneous deamination of methylated cytosine (C) results in the formation of a thymine (T) during the next round of replication if the deamination error is not detected and efficiently corrected by the cell. Since the major proportion of cytosines in CpG dimers are methylated in eukaryotic DNA (van der Ploeg and Flavell, 1980), CpGs may reasonably act as hotspots for C to T transitions.

2. Extent of Polymorphism at Individual Loci

RFLP loci differ significantly in their degree of variability. Most RFLPs characterized to date are two allele polymorphisms (Willard *et al.*, 1985). Since humans are diploid this results in three possible genotypes, two representing the homozygous states, with one type of heterozygote. In general, most DNA polymorphisms are thought to be due to point mutations at restriction enzyme recognition sites (Gusella, 1986). This, by necessity, results in a two allele polymorphism, defined by the presence or absence of the cleavage site.

Although point mutations tend to be the more frequent event, there are multiple levels of variation at individual RFLP loci. Polymorphic insertion or deletion of DNA sequences results in allelic variation which can be quite extensive, ranging from two alleles, to the extreme of

hypervariability, where each individual in the population differs from all other individuals at a given locus. In 1985 Jeffreys *et al.* (1985a), introduced the concept of DNA fingerprints defined by hypervariable minisatellites. Tandem repeats of a core sequence which vary in copy number by an insertion/deletion mechanism, have resulted in complex restriction fragment patterns in the human genome which are heritable, yet unique for each individual. Since then, multiple regions of hypervariability have been identified in the human genome. These include tandem repeat regions associated with the myoglobin gene (Jeffreys *et al.*, 1985a), the zeta-globin pseudogene (Proudfoot *et al.*, 1982), and the insulin gene (Bell *et al*, 1982) loci. These and other basic tandem repeat sequences are thought to exist in as many as 200 copies each per haploid genome (Y. Nakamura, pers. comm.). Several other highly polymorphic loci which display 15 or more fragments, have been identified (Litt *et al.*, 1986; Bufton *et al.*, 1986). Generally, the mechanisms through which these regions arise remain undetermined.

3. Compound Loci

The extent of DNA polymorphism varies regionally as well as at single loci. Clusters of closely linked RFLP loci have been detected and can be treated as single compound loci when close genetic linkage results in the inheritance of the loci as a single unit (Botstein *et al.*, 1980). The information from each RFLP in a compound locus can be combined to form a haplotype. Each haplotype is then

treated as an allele for that locus, substantially increasing the effective heterozygosity.

Although the incidence of close genetic linkage between individual RFLP loci is not known, it is reasonable to assume that it is a common event based upon estimates of numbers of RFLPs in the genome and the average frequency of recombination. Recombination between two markers separated by 1×10^6 bp will occur with an average frequency of 1%. If a RFLP locus occurs every 3000 to 40,000 bp as estimated, the frequency of recombination between neighboring loci would be only 0.003-0.04%. This suggests that closely linked RFLPs occur frequently in the genome.

There has been speculation that the nonrandom association between closely linked DNA polymorphisms would result in linkage disequilibrium, with some of the haplotypes occurring less or more frequently than expected (Chakravarti *et al.*, 1984). Varying degrees of linkage disequilibrium between closely linked loci have been observed, but not with the frequency predicted, and with no consistent relationship between the amount of disequilibrium and physical distance (Litt and Jorde, 1986). In fact, multiple closely linked loci have been detected which show little or no linkage disequilibrium, and are therefore available as highly polymorphic compound loci (Barker *et al.*, 1984; Litt and White, 1985).

4. Measures of the Degree of Polymorphism

The level or degree of variation at a single or compound locus is measured empirically and expressed as a value indicative of the usefulness of that locus as a genetic marker. Individuals which are heterozygous for the alleles of a given marker provide information about segregation at that locus because both chromosomes can be individually detected and followed. In homozygotes, the alleles, and therefore the individual homologues can not be distinguished. Consequently, the frequency of individuals in a population who are heterozygous for a given locus provides an indicator of the value of that locus for genetic linkage analyses. The maximum percent heterozygosity for a locus with two alleles is 50% when the two alleles are present with equal frequency. When the number of alleles is greater than two, as is the case for compound loci and many insertion/deletion polymorphisms, the frequency of heterozygosity in the population can vary from a small proportion to 100%, depending on the allele frequencies. In general, as the number of alleles at a given locus increases, so does the percent heterozygosity and the usefulness of the locus. However, if one allele is predominant, the frequency of heterozygosity drops.

In 1980 Botstein *et al.* introduced the concept of polymorphism information content (PIC) as an alternative means of numerically expressing the usefulness of a polymorphic locus to linkage analyses. The PIC of a locus

is defined as the probability that a child picked at random is informative at that locus. This indicates that informative alleles are segregating in the majority of families, at a frequency where the incidence of heterozygosity is high. PIC values approaching the maximum value of 1.0 indicate that the locus will provide linkage information for virtually every family tested. As with percent heterozygosity, PIC values generally increase with the numbers of alleles at a locus. A two allele polymorphism will have a maximum PIC value of 0.375. The hypervariable minisatellites have PIC values approaching 1.0 (Jeffreys *et al*, 1985).

PIC is based on the number and frequency of alleles for a single locus, or haplotypes for a compound locus, and is calculated as:

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

where n is the number of alleles and p_i and p_j are the allele frequencies.

D. Detection of RFLPs

1. Southern Blot Analysis

The most straightforward and commonly used approach to detect DNA polymorphisms in the genome is screening by Southern blot analysis (Southern, 1975). This technique allows for the direct detection of varying size restriction

fragments in genomic DNA cleaved by a given restriction enzyme. Following restriction digestion of genomic DNA, the fragments are separated according to size by agarose gel electrophoresis. The DNA is transferred by Southern blotting to a solid support such as a nylon membrane, and is then hybridized to a radioactively labeled cloned probe specific for the locus in question. Autoradiography of the hybridized fragments provides visualization of the resulting RFLP pattern.

2. RFLP Probes

Cloned probes which detect RFLPs are recombinant DNA (or RNA) molecules which are homologous to single or low copy number regions of the genome associated with sequence variations in the population. These probes may be specific for restriction site polymorphisms, insertion/deletion polymorphisms or both. A probe may detect a single RFLP locus or multiple closely linked loci. Generally, the likelihood that a given probe will detect more than one variable locus is related to the size of the probe, commonly measured in kilobase pairs (kb).

Until recently, the most common source of cloned probes which detect RFLPs has been from the isolation and cloning of genes of known function and origin. The first human RFLPs were fortuitously detected at the globin gene loci during characterization of those sequences (Maniatis *et al.*, 1978; Jeffreys, 1979). Since then, at least one polymorphic

gene has been detected on each chromosome except the Y (Willard *et al.*, 1985), (indicating that all men are alike).

As the usefulness of probes which detect RFLPs became increasingly recognized, so did the realization that the numbers and distribution of gene-specific probes were limited to coding sequences. To take full advantage of the theoretical ability of RFLPs to exist in all regions of the genome, sequences of unknown function also need to be examined. This second class of RFLP probes, anonymous DNA probes, derive from arbitrary genomic sequences screened at random for RFLPs. Cloned DNA molecules are selected at random from libraries of recombinant human genomic DNAs. These probes are screened by Southern blot analysis for their ability to detect variable single copy fragments on genomic blots. The first RFLP detected by an anonymous DNA probe, pAW-101, was reported by Wyman and White in 1980. Subsequent to this report, the number of RFLPs detected by anonymous DNA probes has increased steadily, reaching approximately 250 by 1985 (Willard *et al.*, 1985).

Probes are capable of detecting RFLPs on Southern blot analysis only if they hybridize to DNA sequences present in the genome at a sufficiently low copy number to be detected as discrete bands. Usually, this is limited to DNA probes which are homologous only to single copy sequences. Some repetitive probes, which have homology to sequences repeated at low levels in the genome also reveal RFLPs. The hypervariable minisatellite probes (Jeffreys *et al.*, 1985a)

and the alpha satellite probes (Jabs *et al.*, 1984; Willard *et al.*, 1985) are examples of this third class of probes which reveal DNA polymorphisms.

E. RFLP Applications

1. Genetic Markers

In 1980, Botstein *et al.*, conceptualized the use of RFLPs as marker loci for the human genome. Theoretically, RFLPs can be used to saturate the genome as genetically linked markers because they are not bound by the constraints of hidden alleles and limitation to coding sequences. The availability of characterized polymorphic loci would make it possible to map other genetic traits to specific regions of the human genome without prior knowledge of their biochemical basis. Close linkage between a RFLP marker and a genetic trait makes it possible to follow the inheritance of the gene in question by visualizing the cosegregation of the phenotype and the marker locus. The use of RFLPs for this purpose is particularly advantageous because of the ease with which RFLPs are detected, and the need for only a small amount of peripheral blood as a source of total genomic DNA for any individual. RFLPs have an additional advantage over many other types of polymorphisms because of the possibility of expansion to include additional linked loci. This creates haplotypes and increases the information potential of the marker.

2. Disease Detection

The availability of RFLPs makes it possible to detect certain genetic disorders of unknown etiology through close linkage to characterized marker loci. Kan and Dozy (1978) presented the first association of a RFLP to a genetic disorder when they demonstrated the coinheritance of a variant Hpa I fragment with the hemoglobin S allele. Prenatal diagnosis of sickle-cell trait confirmed the value of this technique. Since then, verified close genetic linkage between some disease loci and RFLPs has made it possible to diagnose the presence of a disorder without direct access to the defective gene or knowledge of the biochemical defect. In addition, the advantage of early detection through relatively non-invasive means is critical to the management of genetic disease.

Assessment of carrier status through linked RFLPs provides high genetic risk families with information often unavailable through other means. Cystic fibrosis, hemophilia, and thalassemia and phenylketonuria are examples of at least 11 genetic disorders that have linked or gene-associated RFLPs proven effective in antenatal diagnosis (Francomano and Kazazian, 1986).

In addition to prenatal diagnosis and carrier detection, RFLPs have proven useful in the preclinical diagnosis of late onset genetic disorders. Adult onset polycystic kidney disease can now be diagnosed presymptomatically through the use of a linked RFLP,

allowing for early intervention (Reeders *et al.*, 1985). In 1983, Gusella *et al.*, announced the discovery of a RFLP marker linked to Huntington's disease. This provides the option of access to preclinical information to individuals at risk for Huntington's disease, who previously had to wait for clinical symptoms to appear.

When linked RFLPs are used in the diagnosis of genetic disorders, family studies are essential to determine the linkage phase of the family in question. The proper association between the disease trait and its cosegregating allele must be established for correct diagnosis to be made. In general, this requires the availability of blood samples from three generations to perform the appropriate family studies. The family must also be informative for linkage, requiring critical individuals in the pedigree to be heterozygous at the marker locus. Consequently, it is desirable to have highly polymorphic, or multiple closely linked RFLPs for diagnosis of disease.

3. Detection of Genetic Contribution to Disease

Many human diseases are multifactorial, *i.e.* they have both genetic and environmental factors contributing to their onset. The most prevalent of these diseases is cancer. Malignancies are generally divided into those which are familial, and those considered to be sporadic. RFLPs as genetic markers have played significant roles in the advances towards understanding the genetic contribution

towards the development of malignant disease. Probes detecting RFLPs near the retinoblastoma/osteosarcoma gene locus were critical to the elucidation of the mechanism by which loss of heterozygosity at that locus is a crucial step in tumor formation (Cavenee *et al.*, 1983). Understanding this phenomenon has recently resulted in the molecular cloning of that gene (Friend *et al.*, 1986). A similar loss of genetic material has been observed in several other forms of malignancy which may involve a common pathogenetic mechanism (Fearon *et al.*, 1985; Koufos *et al.*, 1985), or additional mechanisms displaying similar behavior (Dracopoli *et al.*, 1985). RFLPs have also been used to determine the clonal origin of tumors (Vogelstein *et al.*, 1985), to define the specific chromosomal breakpoints involved in cytogenetic rearrangements associated with oncogene activation and subsequent malignancy (Rovigatii *et al.*, 1986; Larson *et al*, 1986), and to monitor the involvement of oncogenes in these rearrangements (Klein and Klein, 1985). Additionally, the presence of RFLPs has been used to assist in diagnosis and to monitor the progression of disease in some types of cancer (Krontiris *et al.*, 1985).

The existence of genetic contributions to other multifactorial diseases was suspected prior to the availability of any informative technology. Many diseases not strictly inherited have been long recognized as having strong familial associations. For some of these disorders, strongly implicated candidate genes have been confirmed as

the primary defect. For the many other multifactorial diseases the suggestion of a heritable component is the only clue. Recently, linkage of RFLPs to two separate multifactorial disorders of unknown etiology have provided chromosomal locations for involved genes. This is the first step towards understanding these complex diseases. In the case of familial Alzheimer's disease, the mapping of a linked locus immediately led to the discovery of a candidate gene (St George-Hyslop *et al.*, 1987; Tanzi *et al.*, 1987; Goldgaber *et al.*, 1987). In the second disease, Egeland *et al.*, (1987) demonstrated genetic linkage of bipolar affective disorder (manic-depressive disease) to a set of RFLPs on the short arm of chromosome 11. The cosegregation of DNA markers derived from 11p and clinically diagnosed disease in an Old Order Amish kindred is the first molecular demonstration of genetic contribution to a psychiatric disorder. In additional studies, 11p has been excluded for linkage to bipolar affective disorder in some non-Amish families (Detera-Wadleigh *et al.*, 1987). This suggests that there is genetic heterogeneity involved, with more than one defective gene being capable of producing the clinical entity defined as manic-depressive disease.

Although the markers linked to familial Alzheimer's disease and bipolar affective disorder are not yet useful for preclinical detection of disease susceptibility because of a lack of knowledge of additional causative factors, they

do provide the first molecular tools available to the investigation of these diseases.

4. RFLPs as Tools of Molecular Biology

The mapping of genetic loci has applications for basic molecular biology as well as clinical utility. These areas frequently overlap, as seen in the investigations of multifactorial disorders. The mapping of a gene of unknown product and function to its chromosomal location is a necessary first step towards the cloning and identification of that gene. Several strategies involving the use of RFLPs have evolved to facilitate the cloning of mapped genes.

Chromosome walking techniques have been used to bring researchers closer to genes of interest, with RFLP probes defining the path by acting as genetically and physically linked marker loci along the chromosome (Hadfield, 1983). These methods have been recently expanded to include the physical linking of marker loci separated by very large distances through the use of pulse field gel electrophoresis (PFGE) (Anand, 1986). Restriction enzymes that cut only rarely in human DNA are used to produce large fragments which are separated according to molecular weight by PFGE. The large fragments are then used to determine physical linkages between marker loci that are several hundred kb apart. These techniques have been used to produce a long range physical map of the Duchenne muscular dystrophy gene locus, which is too large to be physically mapped by conventional means (van Ommen *et al.*, 1986).

Another recently developed technique, chromosome jumping, enables researchers to move rapidly from a linked, but distant RFLP, to direct cloning of sequences at or near a gene of interest. The construction of large insert libraries will make it possible to clone DNA sequences up to 100 kb from an initial marker locus. Jumping libraries of this nature have been constructed to facilitate the cloning of sequences at or near the cystic fibrosis gene locus by using the linked *met* oncogene as a starting point (Collins *et al.*, 1987).

The majority of genes mapped to their chromosomal location by association with RFLPs are cryptic genes, or genes of unknown function, generally identified by an associated disease. This provides a substantial challenge in cloning as there is no established means of recognizing the gene. The process of approaching the gene from the DNA rather than the protein level is known as reverse genetics (Orkin, 1987). The single most critical factor in three recent successes using reverse genetics was the molecular detection of deletions at the gene locus of interest. The availability of RFLP probes deriving from the deleted and flanking regions was critical to the identification of the deletions and delineation of their boundaries. Cloned segments were isolated from the sequences deleted in patients with the diseases, and were subsequently identified as to their origin by hybridization with the distinguishing RFLPs. Through the use of these reverse genetics

strategies, sequences have been cloned from the Duchenne muscular dystrophy, retinoblastoma/osteosarcoma, and chronic granulomatous gene loci (Monaco *et al.*, 1986; Friend *et al.*, 1986; Royer-Pokora *et al.*, 1986).

RFLPs have also been useful in evolutionary studies, particularly since they, themselves have arisen as mutations. Sequence variation between species and between human populations has been used extensively to study the rates and paths of evolution (Cavalli-Sforza *et al.*, 1987). The hypervariable α-satellite polymorphisms present at the centromeres of all autosomes and the X chromosome have been used to speculate about the evolutionary divergence of the human chromosomes (Jabs *et al.*, 1984). In addition, linkage disequilibria between pairs of RFLP loci may provide information regarding evolutionary factors such as rates of mutation and recombination, and genetic drift (Antonarakis *et al.*, 1982; Litt and Jorde, 1986).

5. Additional Applications:

There are many additional applications for RFLP marker loci. RFLP probes have been used as cytogenetic markers to distinguish homologous chromosomes (Stewart *et al.*, 1985; Goodfellow *et al.*, 1985). Consequently, DNA polymorphisms can be used to determine the origin of various chromosomal abnormalities such as monosomies, trisomies, insertions or deletions, and translocations. Detection of the origin of

meiotic nondisjunction resulting in chromosome abnormality may be of clinical or scientific benefit, or both.

RFLPs which display hypervariability have been used in some novel approaches towards identification of individuals or their tissues. The DNA fingerprinting technique of Jeffreys *et al.* (1985b, 1985c), has been used successfully for forensic and paternity testing, tissue typing for transplants, and following the progress of transplanted tissue.

F. Construction of a Human Genetic Linkage Map

1. Benefits

The advantages and applications of RFLPs as genetic markers outlined here support the contention of Botstein *et al.* (1980), that construction of a human genetic linkage map is of scientific and clinical importance. The availability of evenly spaced, closely linked marker loci throughout the human genome would provide a map through which any other detectable genetic trait could be localized by demonstration of linkage to the flanking markers. Localization of a gene or trait facilitates the cloning and identification of the defective and normal locus by identifying a region of DNA for study and nearby RFLP probes as molecular markers. Additionally, the surrounding RFLPs provide a readily available set of probes useful for preclinical detection of associated genetic disease.

The existence of evenly spaced, closely linked markers would also provide the information needed to perform multilocus linkage analysis for any locus in the genome. Interval mapping, the process extending from multilocus linkage analysis, is recognized as the most efficient method of detecting linkage, estimating recombination, and determining gene order from family data (Lathrop *et al.*, 1985, 1986; Thompson, 1984). Interval mapping examines the likelihood that a new locus is closely linked to an interval defined by previously placed marker loci, as opposed to being totally unlinked to that region. This provides a sharp dichotomy, which is easier to prove or disprove than is the likelihood of linkage to a single marker, which must be assessed for recombination fractions ranging from zero (close linkage), to 0.5 (unlinked). As a result, fewer informative meioses are needed to prove or disprove linkage (Lander and Botstein, 1987). Obviously, the larger the number of characterized markers for a given region, the greater the efficiency and accuracy of new locus placement. Until recently, this powerful method has been available for only a few regions of the genome due to the paucity of linked markers. Establishment of a detailed human genetic map would allow interval mapping to be used for placement of all newly mapped loci.

2. Requirements:

A complete linkage map of the human genome requires the placement of characterized RFLPs at evenly spaced

chromosomal locations. Botstein *et al.* (1980), estimated that a recombination fraction of 0.2 is the maximum distance between RFLPs that would be detectable in small family structures. Placement of loci at this distance of separation minimizes the number of RFLPs needed to complete a genetic map and still efficiently provide evidence of linkage. Marker loci placed at 20 cM intervals would supply a marker no greater than 10 cM away from any subsequently placed locus. Linkage should be easily detected at these genetic distances. Assuming that the human genome is 33 Morgans in length and that there is an equivalence between Morgans and recombination fractions (Renwick, 1971), it would take a minimum of 165 evenly spaced RFLPs to provide a linkage map based on these parameters (Botstein *et al.*, 1980). Obviously, random isolation and characterization of RFLPs will require that a much larger number of cloned RFLPs be examined in order to map 165 which are distributed evenly at 20 cM intervals.

RFLPs which display a high degree of polymorphism are of maximum usefulness as genetic map points. In practice, mapping of highly polymorphic markers can be accomplished in two ways. Strategies for the isolation of RFLPs that have PIC values approaching 1.0 can be used to select for highly polymorphic loci. Use of large insert vectors such as cosmids to facilitate detection of RFLPs, and probes specific for tandem repeats associated with hypervariable regions are techniques designed to select for highly

polymorphic marker loci (Litt and White, 1985; Nakamura *et al.*, 1987). Alternatively, selected regions of the genome can be saturated with two allele and other low PIC polymorphisms. In this way, the creation of compound loci can provide markers with increased PIC values as desired.

Efficient construction of a linkage map for the human genome also requires the compilation and examination of large amounts of data collected on RFLPs from many sources. Detection of linkage between RFLPs requires that genotype information from a given family or families be accessed for each marker so that patterns of inheritance may be followed. A large number of three generation families, preferably with large sibships in the third generation are necessary, as not all families will be informative for each RFLP probe and not all offspring will be informative in each situation. To facilitate the mapping process, the *Centre d'Etude du Polymorphisme Humain* (CEPH) has compiled 40 large multigenerational families which will act as an information base for the linkage of RFLPs (Marx, 1985). Availability of DNA from these families allows researchers to generate genotype information for specific RFLPs which will then be directly assessed for linkage to loci from other laboratories in an international cooperative effort.

Another effort to compile and disseminate information about the genetic constitution of the human genome is through the biannual international workshop on human gene mapping (HGM). In this setting, maps of each human

chromosome are updated as to the placement of genes and anonymous loci. Each confirmed locus is assigned a number which reflects the nature of the probe, its chromosome location, degree of repetitiveness of the DNA, and the chronological order in which it was mapped. For example, *D16S1* is the annotation used to designate a DNA probe, from chromosome 16, single copy, and the first locus assigned to that chromosome. This information is then published in a single source to provide ready access.

3. Progress

At this point, nearly 400 polymorphic DNA markers have been mapped to human chromosomes (White *et al.*, 1987). Most of these tend to be clustered in the genome, rather than spread evenly throughout. Complete primary genetic linkage maps, with markers spaced approximately every 2 cMs, are available for chromosomes X, 7, 12, and 13, and regions 6p and 11p (White *et al.*, 1987). Most of these markers are diallelic, suggesting that further saturation of these regions, as well as concentrating on unmapped regions, may be beneficial.

4. Problems

One of the greatest difficulties encountered in the construction of a human genetic linkage map is in the non-random distribution of polymorphic loci detected to date. It is not known if this is a reflection of the actual distribution of polymorphism in the genome, or if there has

been bias in the ascertainment of polymorphisms. Strategies for the region specific detection of RFLPs may be necessary to fill in areas currently devoid of marker loci. Attempts to detect RFLPs in preselected regions of the genome facilitates mapping by the more efficient placement of markers. For example, the development of chromosome specific libraries by the use of somatic cell hybrid cell lines which retain only specific human chromosomes, or by flow sorting of chromosomes according to size, provides a means of targeting specific genomic regions (Gusella *et al.*, 1980).

Concentration of mapping efforts on a specific genomic region may also address another of the problems encountered in linkage map construction. Since most markers are diallelic, it may be most advantageous to create compound loci from closely linked RFLPs whenever possible. These linked markers will most likely be detected if a specific genomic region is systematically mapped. In this way, even areas that have no naturally occurring highly polymorphic loci can benefit from markers with high degrees of heterozygosity.

G. Progress Towards Linkage Maps For Chromosome Regions 11q and 16q

Examination of a map of the human genome reveals that two regions particularly devoid of any genetic markers are the long arms of chromosomes 11 and 16 (11q and 16q). The last report of the committee on human gene mapping (HGM8,

1985) showed that only three polymorphic loci are assigned to the long arm of chromosome 11 and just two are mapped to 16q. Consequently, little genetic information regarding either of these regions is available, resulting in a concomitant difficulty in assigning additional loci to these arms.

The strategies used in this thesis for initiating efforts in mapping chromosomal regions 11q and 16q address many of the general difficulties inherent in the construction of a human genomic linkage map. A chromosome specific library was constructed to target the regions of interest. A somatic cell hybrid cell line provided by T.K. Mohandas (UCLA) was used as a means of isolating these genomic regions. The cell line, CF52, has a translocated human chromosome containing 11q and 16q on a mouse genomic background (Bufton *et al.*, 1986). Approximately one to two percent of the hybrid cell genome is human, originating exclusively from chromosome regions 11q and 16q. Therefore, clones from this library identified as human in origin will provide immediate access to the regions of interest. Human clones are selected by their ability to hybridize to a total human DNA probe, and failure to hybridize with a total mouse DNA probe (Fig. 2). Any isolated probe which reveals an RFLP will be of value because of the paucity of genetic markers in these areas.

The chromosome specific library was constructed in a large capacity lambda replacement vector, EMBL-3, with a

maximum insert size of 20 kb (Frischauf *et al.*, 1983). Conditions for construction of the library were established to optimize the size of fragments inserted into the recombinant clones (see manuscript 1 for details). This allows relatively large sections of the genome to be screened for the presence of RFLPs, facilitating the screening process and increasing the likelihood of detecting closely linked polymorphic loci. The ability of a single clone to reveal multiple RFLPs increases with the size of the insert fragment, suggesting an increased efficiency with the largest possible recombinant clones. However, attempts by our laboratory to use cosmids as cloning vectors for this project have been less successful than with the phage library. Instability of cosmid clones in culture made use of our 11q;16q cosmid library inefficient for this purpose.

A strategy of Litt and White (1985), enables us to use large insert fragments for Southern blot analysis without isolation of single copy regions. With this procedure, the repetitive sequences present in most large genomic fragments are prevented from hybridizing to Southern blots by blocking them with homologous genomic DNA. The single stranded DNA probe is annealed with an excess of total human DNA under conditions which drive the repetitive sequences into duplex form, while the single copy regions remain available for hybridization to the Southern blot. This allows multiple linked single copy regions that are physically separated by repetitive sequences to be screened at one time without

subcloning or further manipulation (see manuscript 1 for details).

In this thesis, initial Southern blot analyses for the detection of clones revealing RFLPs were done under conditions established for efficient screening. Southern blots generated by digesting the DNAs from six unrelated individuals with Taq I, Msp I or Rsa I were screened initially to determine which human clones detect RFLPs. Taq I and Msp I have four bp recognition sequences with internal CpG dinucleotides. Consequently, they cut frequently within the human genome, specifically at sequences which have a higher than average mutation rate (Barker *et al.*, 1984). They should, therefore, be more efficient than other enzymes at detecting RFLPs. Rsa I was selected because empiric data from Collaborative Research Inc. suggests that it reveals RFLPs at a higher than average rate (Donis-Keller *et al.*, HGM8, 1985). Once a clone is selected by its ability to reveal Taq I, Msp I, or Rsa I polymorphisms it is screened against approximately ten other frequently cutting enzymes to detect additional polymorphic loci.

Recombinant probes which have large inserts homologous to multiple single copy regions frequently give complex fragment patterns on Southern blot analysis. If multiple polymorphic loci are detected by a single probe, it can be difficult or impossible to determine the number of loci, and which fragments are allelic. Subcloning the insert as smaller restriction fragments will separate loci so they can

be detected alone by the newly created subset of probes. The simplification of resulting Southern blot patterns usually resolves the dilemma of allelism, and often results in less interference from background noise on the blot. Subcloning is however technically troublesome and time consuming. We are often able to eliminate the need for subcloning by subdividing complex probes using restriction enzyme digestion and fragment separation by agarose gel electrophoresis. Restriction fragments are excised from the gel and radioactively labeled directly by the method of Feinberg and Vogelstein (1983). These fragments are then used directly on Southern blots without further purification of the DNA from the agarose. In this way, segments of the original clone which reveal RFLPs are quickly and easily isolated without subcloning and extensive screening (see manuscript 2), and frequently have an improved signal to noise ratio .

An additional time saving strategy used in this thesis involves the removal of non-specifically hybridizing material from probes. It is not uncommon to find that some DNA probes are prone to producing high background on Southern blot analysis. We have found that a brief mock hybridization of the probe to a Southern blot of no experimental consequence will frequently remove the interfering material from the hybridization mixture, leaving the probe clean and intact for subsequent hybridization to

the desired Southern blot (S. Kondoleon, pers. comm.) (see manuscript 2).

Assessment of linkage between marker loci mapped in this study is carried out through the use of likelihood ratios. Usually genetic linkage analysis involves uncertainty regarding the phase, or specific chromosomal arrangement of genotype, in the parents of the families being analyzed. Since we can not arrange for informative backcrosses in humans to determine the linkage phase, we deal with the uncertainty by assessing linkage on the basis of likelihood ratios (Fisher, 1935; Haldane and Smith, 1947; Morton, 1955). This compares the probability that observed genotypes would arise under the hypothesis of linkage between two loci at a defined interval of recombination (), with the probability that the loci are genetically unlinked. This is known as the *odds ratio*. Odds ratios from independent families can be combined, building evidence for or against linkage by including additional informative meioses. The information is conventionally stated as the log of the odds ratio, known as a LOD score. Data resulting in an odds ratio of at least 1000:1 in favor of linkage (a LOD score of 3.0) is considered to be proof of genetic linkage. Odds of 100:1 against linkage is sufficient to reject that possibility. A maximum LOD score value at a given value of provides evidence for a genetic distance of separation equal to that recombination fraction.

Computer routines make it feasible to analyze large numbers of linkage hypotheses in reasonable periods of time. Although many of the LOD scores presented in this thesis were calculated by hand, the computer program, LINKAGE (Lathrop *et al.*, 1984), was also used for some calculations as designated in the manuscripts.

H. Summary

It is clear that the efficient construction of a human genetic linkage map will require the utilization of novel methodologies which expedite the mapping of RFLPs by overcoming the difficulties inherent in these analyses. This thesis has made use of several unusual tactics to facilitate the mapping of RFLPs targeted to the virtually unmapped regions of 11q and 16q. Combining these strategies in this focused attempt has resulted in the successful cloning and characterization of a novel cluster of at least 15 RFLP loci deriving from chromosome region 16q, and the development of six polymorphic loci which map to the distal portion of 11q. The results of this thesis are summarized in the following two manuscripts and Appendix 1.

Figure 1a. Restriction site polymorphism. Presence or absence of the variable restriction site, indicated by an asterisk, results in varying fragment lengths of 8.0, 5.0, and 3.0 kb.

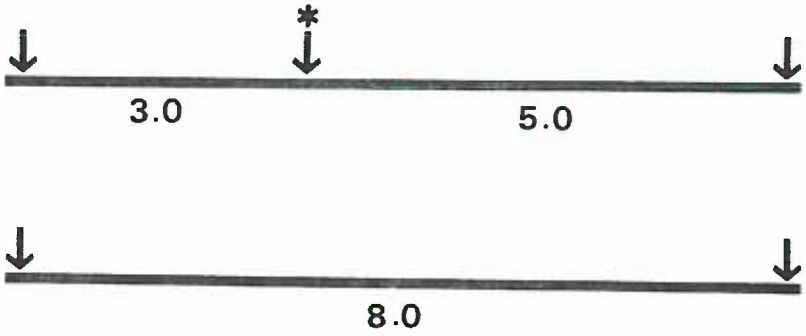


Figure 1b. Insertion/deletion polymorphism. Polymorphic insertion or deletion of a 1 kb fragment between flanking restriction sites results in restriction fragments varying by 1 kb in length.

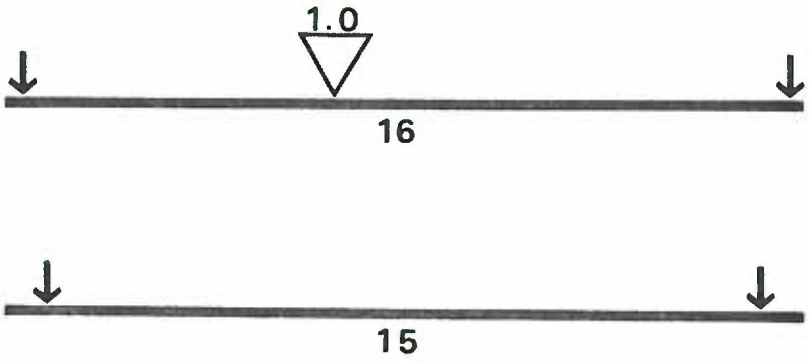
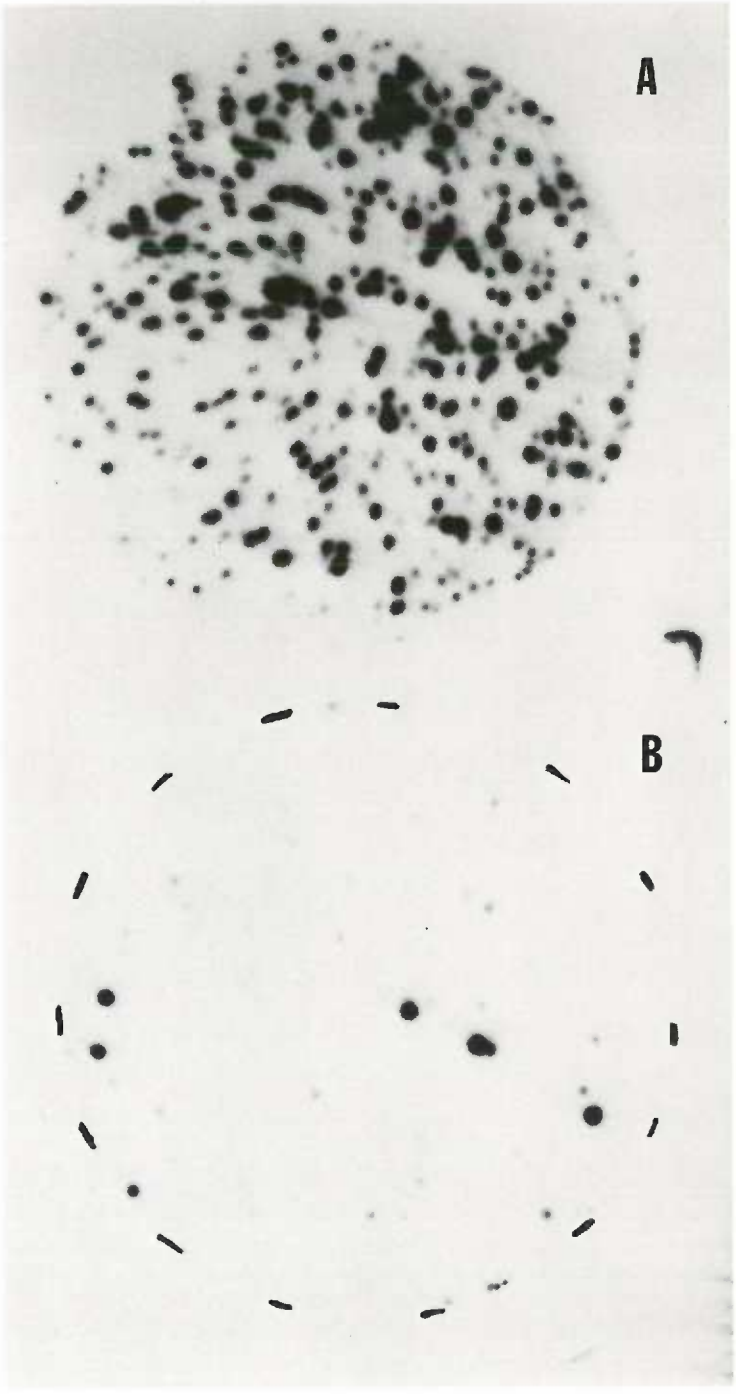


Figure 2. Panel A shows a plaque screening filter hybridized with a radiolabeled probe total mouse DNA. Only recombinant plaques which carry a mouse DNA derived inserts are detected.
Panel B shows the same filter, stripped of the mouse DNA probe, and reprobed with a total human DNA probe. Only the recombinant plaques carrying human DNA derived inserts are detected.


A
B

II. MANUSCRIPT 1:

An Unusual Clustering of Restriction Site Polymorphisms on Chromosome 16q

Summary:

A single recombinant phage,Φ8-9, reveals at least 15 linked polymorphic loci mapping to chromosome 16q22-24. This extensive variation is due to a clustering of individual restriction site polymorphisms rather than resulting from an insertion/deletion mechanism. The range of restriction enzymes (at least 11) that detect these loci suggests that mutation at CpG dinucleotides did not play a predominant role in the formation of this cluster. Consideration of haplotypes constructed from just three of the loci results in a population heterozygosity of 86%.

Introduction:

The development of a human genetic linkage map using restriction fragment length polymorphisms (RFLPs) as markers will provide an important means of localizing any gene of interest in the genome (Botstein *et al.*, 1980). Placement of a locus by linkage analysis should facilitate the eventual isolation, cloning, and identification of these genes. RFLPs have played an important role in the recent cloning of the genes for chronic granulomatous disease (Royer-Pokora *et al.*, 1986), Duchenne muscular dystrophy (Monaco *et al.*, 1986) and retinoblastoma/osteosarcoma (Friend *et al.*, 1986). Additionally, RFLPs have recognized roles in prenatal and preclinical diagnosis of certain genetic disorders such

as cystic fibrosis (Tsui *et al.*, 1985; Wainwright *et al.*, 1985; White *et al.*, 1985), the investigation of mechanisms fundamental to hereditary cancer syndromes (Koufos *et al.*, 1985; Fearon *et al.*, 1985) and nonhereditary malignancies (Vogelstein *et al.*, 1985; Dracopoli *et al.*, 1985). RFLPs have also been fundamental in the identification of genetic factors contributing to a wide variety of multifactorial disorders such as familial Alzheimer's disease and bipolar manic depressive disorder (St George-Hyslop *et al.*, 1987; Tanzi *et al.*, 1987; Egeland *et al.*, 1987).

At least two factors have been identified as important for the efficient construction of a genetic linkage map. Even spacing of marker loci along each chromosome at an estimated 20 cM separation is desirable because it should allow the reliable placement of any other locus through linkage analysis (Botstein *et al.*, 1980). Marker loci with polymorphism information contents (PIC) approaching 1.0 are most useful because the probability that informative alleles are segregating in a given family is at a maximum. Use of high PIC marker loci minimizes the number of families needed to provide sufficient data to establish linkage (Botstein *et al.*,1980). However, most previously reported marker loci are two-allele polymorphisms with concomitantly low PIC values. Only about 10% of reported RFLPs have PIC values greater than 0.5 (Willard *et al.*, 1985). Currently

available marker loci also tend to be clustered in the genome, leaving regions virtually devoid of any map points. Chromosome region 11q has only three assigned marker loci, with APOAI as the single 11q marker with a PIC >0.5. A similar situation exists on 16q, with just two RFLP assignments. The paucity of RFLPs assigned to 11q and 16q led us to construct chromosome specific cosmid and phage libraries from a human-mouse somatic cell hybrid with an 11q;16q translocation as the sole human chromosome. As a result, a third marker for region 16q, provisionally named D16SPDX1, was recently reported from this laboratory (Bufton *et al.*, 1986a). Here we report the detection and characterization of a highly polymorphic region mapping to 16q22-24.

In our approach, whole recombinant phage DNAs are hybridized directly to Southern blots after prehybridization of the radioactively labeled probe with a vast excess of total human DNA under conditions which drive only the repetitive DNA into duplex form (Litt and White, 1985; Sealy *et al.*, 1985). This allows rapid screening of large numbers of clones for their ability to detect RFLPs. The prehybridization technique also makes it possible to map probes by in situ hybridization without the isolation of single copy subclones (Bufton *et al.*, 1986a).

A single phage clone,Φ8-9, was found to detect an unusual cluster of closely linked restriction site

polymorphisms, estimated to encompass at least 15 loci. This series of related loci has been designated D16S20. When just three of these loci are combined to form haplotypes, a genetic marker with a population heterozygosity of 86% results.

Materials and Methods:

Somatic Cell Hybrids:

DNA used to construct the chromosome specific library was obtained from a previously described mouse x human somatic cell hybrid (CF52) that retained a single translocated human chromosome, t(11;16)(q13;p11), on a rodent background (Bufton *et al.*, 1986a).

Cell hybrids for the mapping panel were provided by Gail Bruns and have been previously described (Bufton *et al.*, 1986b).

Phage Library Construction:

The 11q;16q chromosome specific library was constructed using the lambda replacement vector, EMBL-3 (Frischauf *et al.*, 1983). Phage arms were prepared with a Bam HI/Eco RI double digestion followed by isopropanol precipitation to remove the stuffer fragment and prevent religation of stuffer and arms.

High molecular weight DNA from CF52, prepared as described previously (Bufton *et al.*, 1986a), was partially digested with Sau 3A to maximize the proportion

of fragments in the 20 kb range (Maniatis *et al.*, 1982). The Sau 3A fragments were dephosphorylated with calf intestinal alkaline phosphatase according to Boehringer Mannheim Biochemical (BMBiochemica, 1986). The insert DNA was then ligated to phage arms in equal molar amounts. The reaction was catalyzed by T4 DNA ligase at 15^{o}C for 16 hours. Ligated DNA was packaged into phage particles using the Gigapack in vitro packaging system (Stratagene, San Diego, CA). Recombinant phage were selected by their ability to form plaques on the restrictive bacterial host, P2392 (Stratagene, San Diego, CA). Clones containing human inserts were selected by hybridizing replica filters (Millipore HATF 082 nitrocellulose) of plaques with labeled total human or total mouse DNA (Gusella *et al.*, 1980). Total human and mouse DNA probes were prepared by oligolabeling with [^{32}P] dCTP (Feinberg and Vogelstein, 1983). Filter hybridizations were carried out in 50% formamide, 5X SSPE, 5X Denhardt's solution, 0.5 mg/ml sonicated, denatured salmon sperm DNA at 43^{o} C, overnight. Plaques which hybridized with human but not with mouse DNA were isolated from the plate as agar plugs, and stored in SM over chloroform (Maniatis *et al.*, 1982).

RFLP Screening:

Plaque-purified recombinant phage were amplified in the permissive bacterial host, LE392, by preparing 8 ml high titer lysates. For the initial screening, DNA minipreps were prepared from 6 ml of the lysate (Maniatis *et al.*, 1982). Whole phage DNAs were labeled with [^{32}P] dCTP by oligolabeling and, after prehybridization with total human DNA were used to probe Southern blots of Taq I, Msp I, and Rsa I digested DNA from panels of unrelated individuals as previously described (Litt and White, 1985). Briefly, prehybridization of the probe is accomplished by annealing the probe to an excess of sonicated total human DNA (2.5 mg/ml in 0.12 M phosphate buffer) to a C_{ot} value of approximately 100 moles/liter/second. Probes revealing RFLPs were examined on Southern blots of DNAs from unrelated individuals digested with at least ten additional restriction enzymes to characterize further the RFLPs and to detect additional linked polymorphic loci.

DNA Isolation:

DNA from recombinant clones of interest was prepared by further amplification of the phage, and subsequent phage and DNA isolation according to Silhavy *et al.* (1984), with final purification by two cesium chloride/ethidium bromide centrifugations.

Linkage Analysis:

Lymphoblast cell lines from 9 three-generation Utah reference families, each with all four grandparents and at least six children present, were obtained from the Institute for Medical Research, Camden, NJ.

Data from examining the Utah reference families was used to determine linkage relationships between marker loci. When linkage phase could not be ascertained, calculation of recombination frequencies was based on the mean of each of the two possible allele distributions in the parents. In those cases where it was possible to determine linkage phase by examining grandparental genotypes, recombinant and nonrecombinant children were identified directly. LOD scores were calculated using both phase known and phase unknown data, at a recombination fraction of zero (Morton, 1962).

Five of the diallelic loci were examined for linkage disequilibrium in all possible pairwise combinations. Chi-square (X^2) values were calculated for goodness of fit for observed versus expected haplotype frequencies. The associated probabilities, p, were calculated for one degree of freedom.

Chromosomal Localization:

Chromosomal localization was achieved by two independent methods: use of a rodent x human somatic cell hybrid mapping panel and *in situ* hybridization. The

hybrid cell panel and the methods used for *in situ* hybridization have been previously described (Litt *et al.*, 1986; Bufton *et al.*, 1986b).

The mapping panel blot was probed with the cloned insert prepared from Sal I digested phage DNA. The digested DNA was run on low gelling temperature agarose and the insert fragment was cut out of the gel. The 17.3 kb insert DNA was then oligolabeled directly in the agarose, prehybridized with total human DNA, and hybridized to the blots.

Probe for *in situ* hybridization was also prepared from the phage insert fragment. Insert DNA was labeled by a modification of the oligolabeling procedure of Feinberg and Vogelstein (1983), with [^{3}H] dCTP and dTTP without removal of the agarose, to a specific activity of >10^{8} dpm/g (T. Donlon, pers. comm.). The labeled DNA was then purified by solubilizing the agarose in an equal volume of 50 mM Tris pH 7.5/0.1% SDS/10 mM EDTA/ 0.5 M NaCl at 68^{o}C for 10 minutes. Excess sonicated total human DNA was added as a carrier for the purification and for the subsequent prehybridization reaction. The DNA was extracted twice with 37^{o}C phenol, twice with chloroform, and precipitated with ethanol in the presence of 2.5 M ammonium acetate. The final DNA pellet was resuspended in 1 M phosphate buffer, and the probe prehybridized with the total human DNA.

Results:

A phage library specific for human chromosome arms 11q and 16q was constructed from a somatic cell hybrid that retained an 11q;16q translocation as the only human chromosome. A recombinant phage,Φ8-9, was isolated from this library and characterized for its ability to detect RFLPs. After prehybridization with total human DNA to block repetitive sequences, this single phage with a 17.3 kb human DNA insert, detects the eight closely linked polymorphic loci seen in Table 1. At least seven additional loci are seen with Taq I, Pst I, Bcl I, Asp 700, and Hae III (data not shown). The complex fragment patterns seen with these loci have so far prevented determination of allelism.

Eleven of the restriction enzymes examined reveal RFLPs with this probe. Nine of these enzymes do not have CpG dinucleotides in their recognition sequences. Eight additional enzymes show no polymorphisms. For the six loci examined (A, B, C, D, E and F), the distribution of genotypes in unrelated individuals indicates that the polymorphisms result from mutations at individual restriction sites within the closely linked loci, and not from insertion/deletion events. This interpretation is supported by the failure ofΦ8-9 to reveal RFLPs with Eco RI, Bam HI, HindIII, Rsa I, HinfI, Xba I, Apa I, and Dra I. With each of the enzymes Sac I and Msp I the probe reveals two RFLPs. In these cases, assignment of allelic

fragments to specific loci was made by observing the segregation of the fragments in several Utah reference families.

Population Studies

Three of the diallelic loci revealed byΦ8-9 were selected for detailed analysis. Loci A and B are detected with Sac I (Fig. 1a). Locus A has allelic fragments of 10.0 kb, and cosegregant 7.4 + 2.6 kb bands. Locus B consists of a 2.9 kb allele, and a second allele seen as two bands of 1.9 and 1.0 kb. Individuals display a maximum of six variable bands, in the case of a double heterozygote, with a minimum of two polymorphic fragments seen in a double homozygote. Locus D, detected with Pvu II, has a 6.5 kb allele, with the 5.8 plus 0.7 kb bands representing the second allele (Fig. 1b). The associated allele frequencies for the three loci are included in Table 1. Observed and expected genotype frequencies for all three loci, based on 36 unrelated individuals tested for each locus, are in good agreement, indicating that each locus is in Hardy-Weinberg equilibrium. Of 36 unrelated Caucasians, 31 were heterozygous at one or more of these three loci. Analysis of the resulting haplotypes of these individuals indicates that all possible haplotypes for the A and B loci are represented at expected frequencies in the population studied. When haplotype frequencies for the A and D, and the B and D loci are examined, significant linkage disequilibrium is

observed. These data and further information suggesting varying degrees of linkage disequilibrium between additional loci are summarized in Table 2.

Three locus haplotypes were determined for 24 of the initial population of 36 unrelated individuals, including 17 grandparents from the Utah reference families. Haplotypes were inferred for five of the grandparents from known linkage phases derived from three generations of family data. Haplotypes could not be derived for any of the five triply heterozygous, nor seven of 12 doubly heterozygous individuals from the original population of 36. The resulting population heterozygosity was 79%, considerably lower than the original random population heterozygosity of 86%.

Linkage Analyses

Detection of multiple polymorphic loci by a single phage clone suggests, but does not prove close linkage of the loci. For D16S20, close linkage is supported by physical evidence of colinearity of the phage insert and a single segment of the genome, and by genetic evidence through linkage analysis. The probe is shown to be colinear with the genome by its hybridization to a discrete number of nonpolymorphic genomic restriction fragments equal to the number of insert-derived fragments generated by that restriction enzyme. Thus, Eco RI digestion of Φ8-9 produces three insert-derived fragments

of 11.5, 4.0 and 1.8 kb. Consequently, if the probe insert is colinear with the genome, it should hybridize to a maximum of three genomic fragments of comparable or larger size. In fact, Southern blot analysis of Eco RI digested total human DNA reveals three genomic fragments of >20, 12.5 and 1.8 kb. Similar results were seen with Bam HI which produces insert-derived fragments of 8.3, 5.5 and 3.5 kb. Southern blots prepared with Bam HI and probed withΦ8-9 reveal three bands of 13.0, 5.5 and 4.9 kb.

Southern blot analyses of five large, three generation Utah kindreds were used to confirm close genetic linkage of loci A, B, C, and D. LOD scores resulting from these analyses are presented in Table 3. No recombinants have been detected in theΦ8-9 cluster. Representative genotypes of families are shown in Southern blots prepared from Sac I and Pvu II digested DNA (Fig.1). In each case, codominant Mendelian inheritance is displayed.

Chromosomal Localization

The insert fromΦ8-9 was used to probe a Southern blot containing HindIII digested DNA from 25 cell lines of a somatic cell hybrid panel. The hybridization pattern depicted as the presence (+) or absence (-) of hybridization is summarized in Table 4. All 11 of the cell lines which contain chromosome 16 hybridized with

the probe. All other chromosomes are discordant with at least 5 cell lines.

A more precise localization was obtained by *in situ* hybridization. A representative, sequentially stained metaphase spread from the *in situ* hybridization study is shown in Fig. 2. The silver grains revealing hybridization of the probe are observed with standard stain while the chromosomes to which the probe has annealed are identified by fluorescent R-banding. Fig. 3 summarizes the results of scoring 110 labeled metaphase spreads from a normal male. 21 of the 110 metaphases (19%) showed hybridization to the distal portion of 16q, (16q22-24); the remainder of the label was distributed randomly over the chromosomes.

Discussion:

We have detected an unusual clustering of restriction site polymorphisms mapping to chromosomal region 16q22-24. The clone which detects this cluster was derived from a genomic phage library specific for human chromosome arms 11q and 16q. The recombinant phage Φ8-9 detects at least 15 polymorphic loci revealed by 11 restriction enzymes. The estimate of 15 loci may be conservative since some restriction enzymes, such as Sac I and Msp I, detect multiple loci. We have found no evidence that an insertion/deletion mechanism played a role in the formation of any of the RFLPs in the cluster.

All loci examined appear to be products of mutations at individual restriction enzyme cleavage sites within the locus. The wide range of restriction enzymes which reveal polymorphisms in conjunction with probeΦ8-9 suggests that the CpG dinucleotide, previously implicated as a common site of mutation in the human genome (Barker *et al.*, 1984), did not play a predominant role in the formation of this unusual cluster of restriction site polymorphisms. Only 2 of the 11 enzymes found to reveal RFLPs with this probe have a CpG dinucleotide in their recognition sequence.

Detection of multiple polymorphic loci with a single recombinant phage suggests close linkage among those loci. Restriction enzyme and Southern blot analyses indicate that all of the loci detected byΦ8-9 are, in fact, closely linked. In genetic linkage analysis of four loci, LOD scores exceed 3.0 in two cases. Although other LOD scores are lower due to the unavailability of informative families, no recombinants were detected.

We have combined three of the loci (A, B, and D) to create a compound locus, characterized by eight haplotypes with a heterozygosity of 86%. No significant linkage disequilibrium was detected between A and B. A lack of linkage disequilibrium between some closely linked loci has been observed previously (Litt and Jorde, 1986; Barker *et al.*, 1985). A significant amount of

linkage disequilibrium has been detected between the A and D, and the B and D loci. This reduces the observed number of heterozygotes from the number expected if no disequilibrium existed. Despite this disequilibrium, the heterozygosity for the population studied is sufficiently high at 86% to provide a very useful locus for gene mapping. In addition, some individuals triply homozygous at this existing compound locus are heterozygous at other closely linked loci, indicating that the heterozygosity will increase if additional loci within the cluster are typed.

The PIC value calculated for the compound locus described here proved misleading. Since this calculation required determination of three locus haplotypes significant bias was introduced by eliminating a substantial number of double and triple heterozygotes. This effectively reduced the population heterozygosity from 86% to 79%. The PIC value calculated from this data reflects that bias and underestimates the usefulness of this locus. For highly polymorphic compound loci composed of multiple closely linked RFLPs, population heterozygosity is more suitable than PIC as an indicator of information content.

Only three other polymorphic loci have been mapped to 16q. Two of these, adenine phosphoribosyl transferase (APRT) at 16q24 and haptoglobin, at 16q22.1 (Fratini *et al.*, 1986; Willard *et al.*, 1985), have not been evaluated

for PIC or heterozygosity. The third polymorphic locus, D16S7, has a PIC of 0.77 and maps to 16q22-24 (Bufton *et al.*, 1986a). The discovery of the highly polymorphic region described here adds a valuable genetic marker to the sparsely populated 16q region.

Several other clones from our 11q;16q specific library are being examined for their ability to detect polymorphic loci mapping to arms 16q or 11q. Probes which reveal RFLPs will be used to type the CEPH families (Marx, 1985), facilitating construction of linkage maps of these human chromosomal regions.

Acknowledgments. Excellent technical assistance was provided by Ruth Litt, Miyoung Kwak and Gwen Hansen. We thank Gail Bruns for providing the somatic cell hybrid mapping panel. We also thank Sophia Kondoleon and Elaine Ostrander for critical comments on the manuscript. This work was supported by NIH Grant RO1-GM32500, Maternal and Child Health Grant MCT920, and the Retinitis Pigmentosum Foundation Fighting Blindness. Cheryl Maslen was supported in part by a Research Fellowship from the N.L. Tartar Trust Fund.

Table 1

RESTRICTION ENZYMES REVEALING RFLPs WITHΦ8-9

Enzyme	Constant Bands (kb)	Polymorphic Bands Allele	Lengths (kb)	Frequency	N[a]
Sac I	4.1, 2.6, 2.1, 1.3	A1,A2	10.0, 7.4 + 2.6	0.43, 0.57	76
		B1,B2	2.9, 1.9 + 1.0	0.26, 0.74	76
Pvu II	3.1, 2.7, 2.6, 1.9, 1.4, 1.2, 1.0, 0.4	D1,D2	6.5, 5.8 + 0.7	0.32, 0.68	114
Bgl II	8.0	C1,C2,C3	>20, 13.7, 8.5 + 5.2	0.03,0.71,0.26	34
Bgl I	2.1, 1.9, 1.7, 1.0, 0.90, 0.87, 0.70	E1,E2	13.0, 11.0 (3.3)[b]	0.71, 0.29	34
Msp I	2.0, 1.4, 1.3, 1.0	F1,F2	3.5, 0.80	0.38, 0.62	16
		G1,G2	1.5, 1.20	0.50, 0.50	16
Bst NI	0.62	H1,H2	0.46, 0.42	0.44, 0.56	12

[a]N is the number of chromosomes in the sample.

[b]Number in parentheses represents a rare allele.

Table 2

PAIRWISE LINKAGE EQUILIBRIUM ANALYSES

Loci	X^2	N[a]	p[b]
A-B	1.50	62	0.25
A-D	10.65	62	0.00
A-E	0.43	24	0.50
B-D	6.85	58	0.01
B-E	2.42	22	0.13
C-E	2.15	14	0.15
D-E	2.16	24	0.15

[a] N is the number of chromosomes.

[b] p is the probability calculated for 1 degree of freedom.

Table 3

GENETIC LINKAGE ANALYSES FOR LOCI IN D16S20

Linkage Group	Recombination Fraction	LOD Score	Number of Informative Meioses
A-B	0.0	2.4	9
A-C	-	-	0
A-D	0.0	2.4	9
B-C	0.0	1.8	7
B-D	0.0	3.9	14
C-D	0.0	3.3	12

Table 4

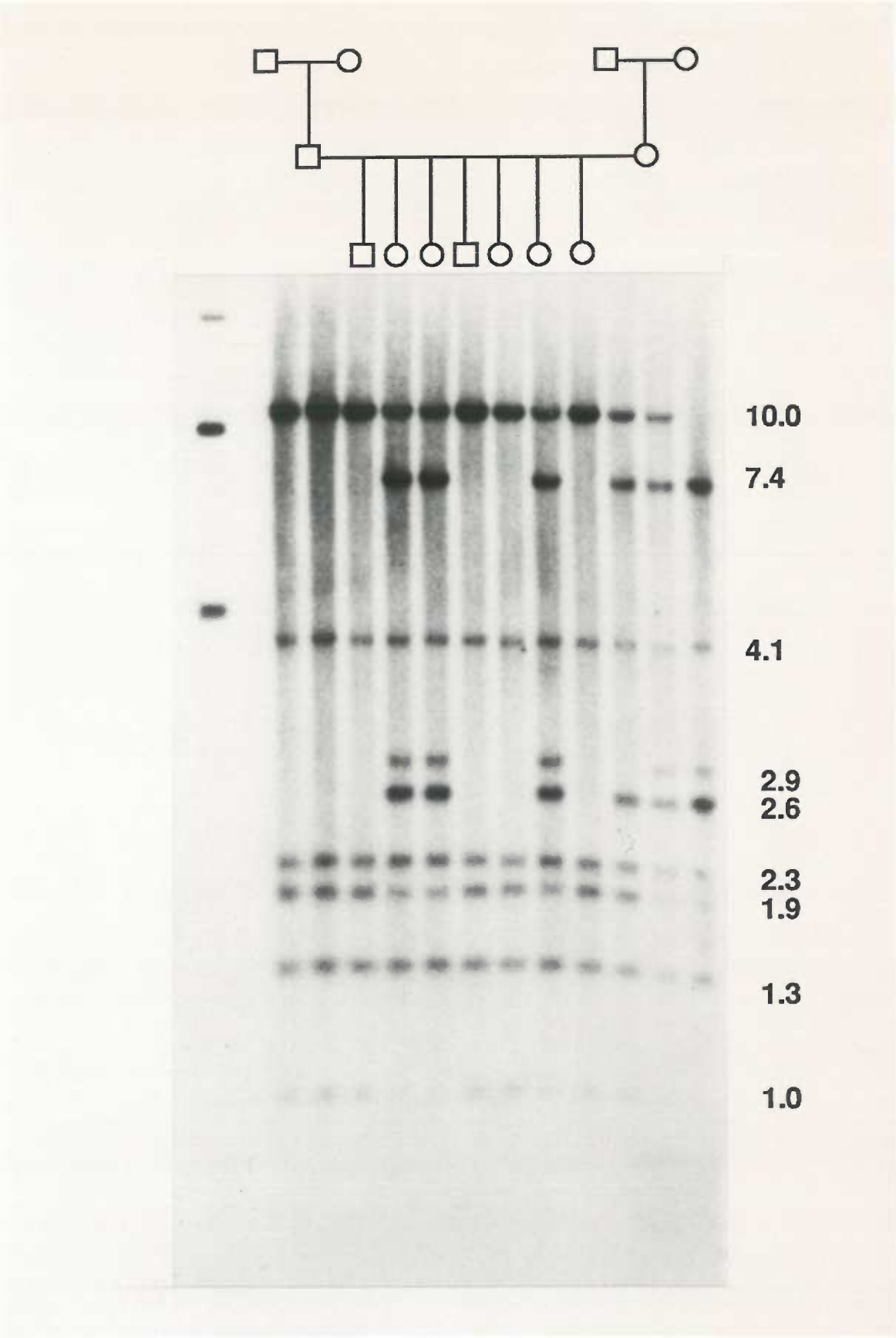
SUMMARY OF MAPPING PANEL

Chromosome	Hybridization Pattern +/+	-/-	+/-	-/+	Discordant Fraction
1	4	8	2	5	0.37
2	3	11	1	6	0.33
3	6	11	1	4	0.23
4	5	6	4	5	0.45
5	5	8	4	6	0.43
6	6	6	5	5	0.45
7	7	10	2	4	0.26
8	6	9	3	5	0.35
9	4	8	2	6	0.40
10	6	7	4	5	0.41
11	5	8	4	4	0.38
12	4	6	4	6	0.50
13	5	8	2	6	0.38
14	7	6	6	3	0.41
15	4	7	4	8	0.52
16	11	12	0	0	0.00
17	5	9	2	6	0.36
18	4	6	6	6	0.55
19	9	2	8	1	0.45
20	6	5	6	4	0.48
21	8	7	4	4	0.35
22	6	8	4	5	0.39
X	3	0	5	0	0.62
Y	1	11	0	10	0.45

The column designations are: +/+, hybridization signal and chromosome both present; -/-, hybridization signal and chromosome both absent; +/- hybridization signal present but chromosome absent; -/+, hybridization signal absent but chromosome present.

Figure 1a. Southern blot of Sac I digested DNAs from Utah reference family K1341 probed with phage Φ8-9. The left lane contains size standards.

Figure 1b. Southern blot of Pvu II digested DNAs from Utah reference family K1345 probed with phage Φ8-9. The right and left flanking lanes contain size standards.


10.0
7.4
4.1
2.9
2.6
2.3
1.9
1.3
1.0

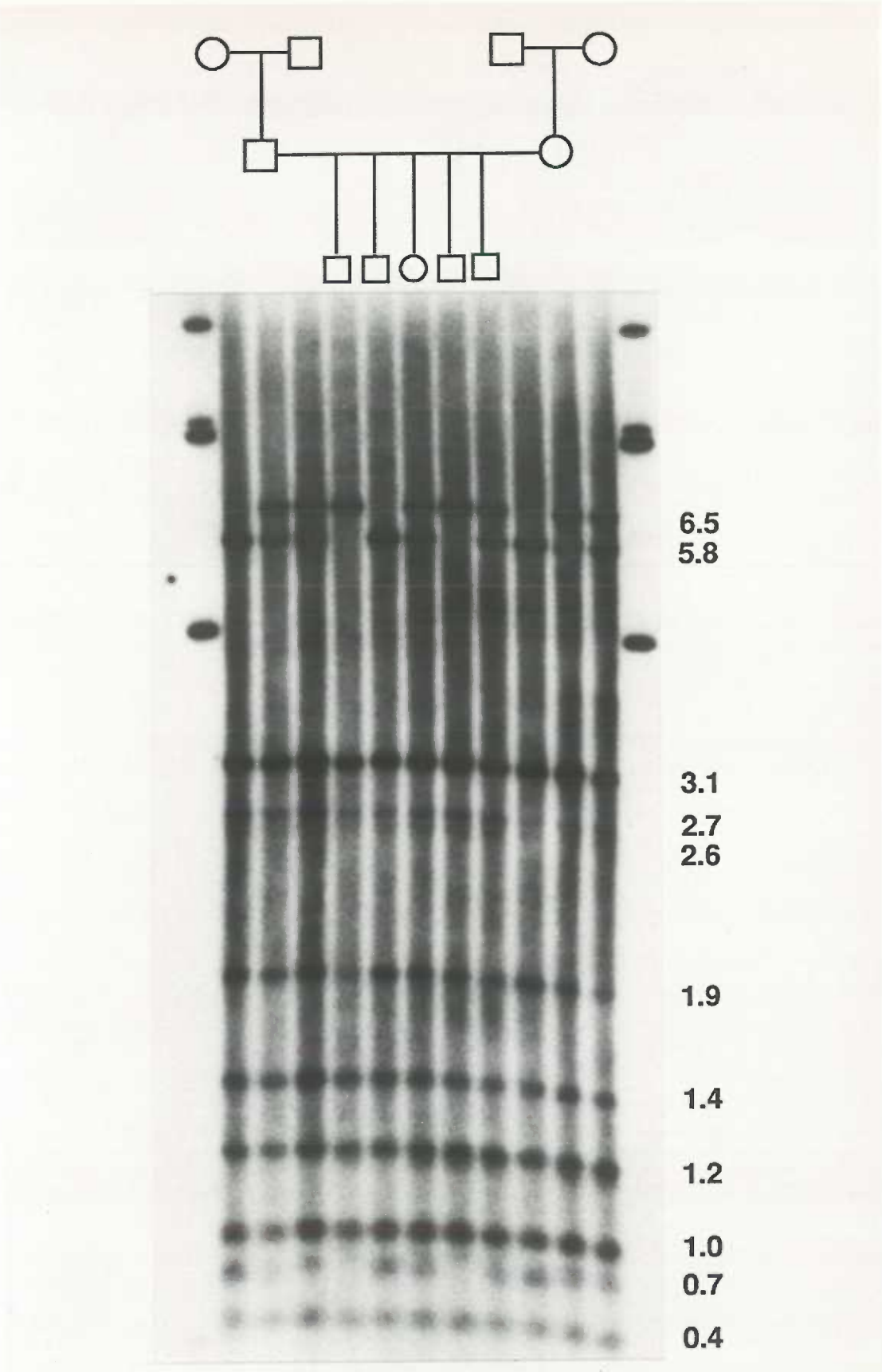

6.5
5.8
3.1
2.7
2.6
1.9
1.4
1.2
1.0
0.7
0.4

Figure 2. *In situ* hybridization of phageΦ8-9. Upper panel:
Representative metaphase that has been R-banded following hybridization.
Lower panel: the same metaphase, destained and restained with Wright's stain to show silver grain located over the tip of 16q and the corresponding location in the R-banded metaphase.

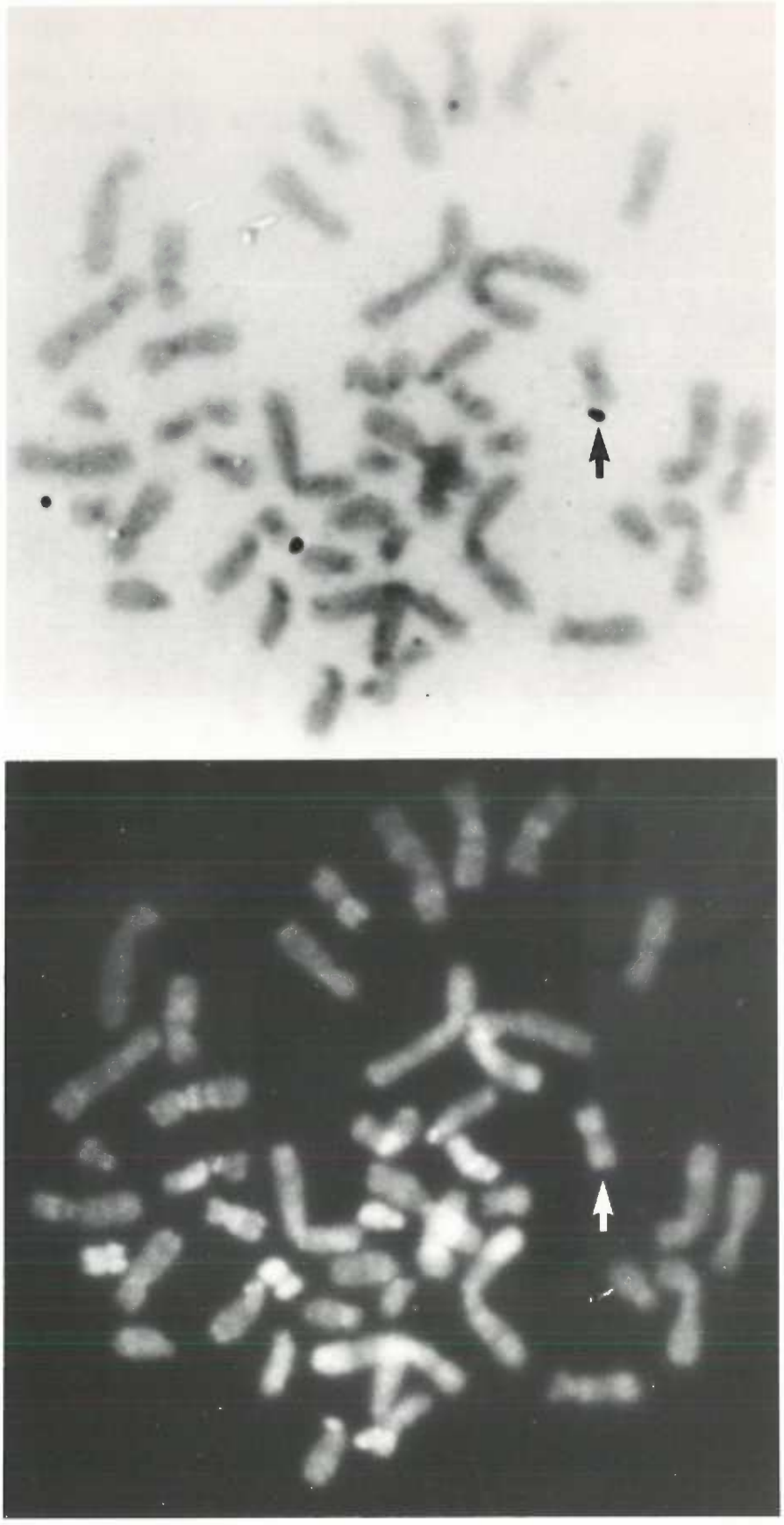
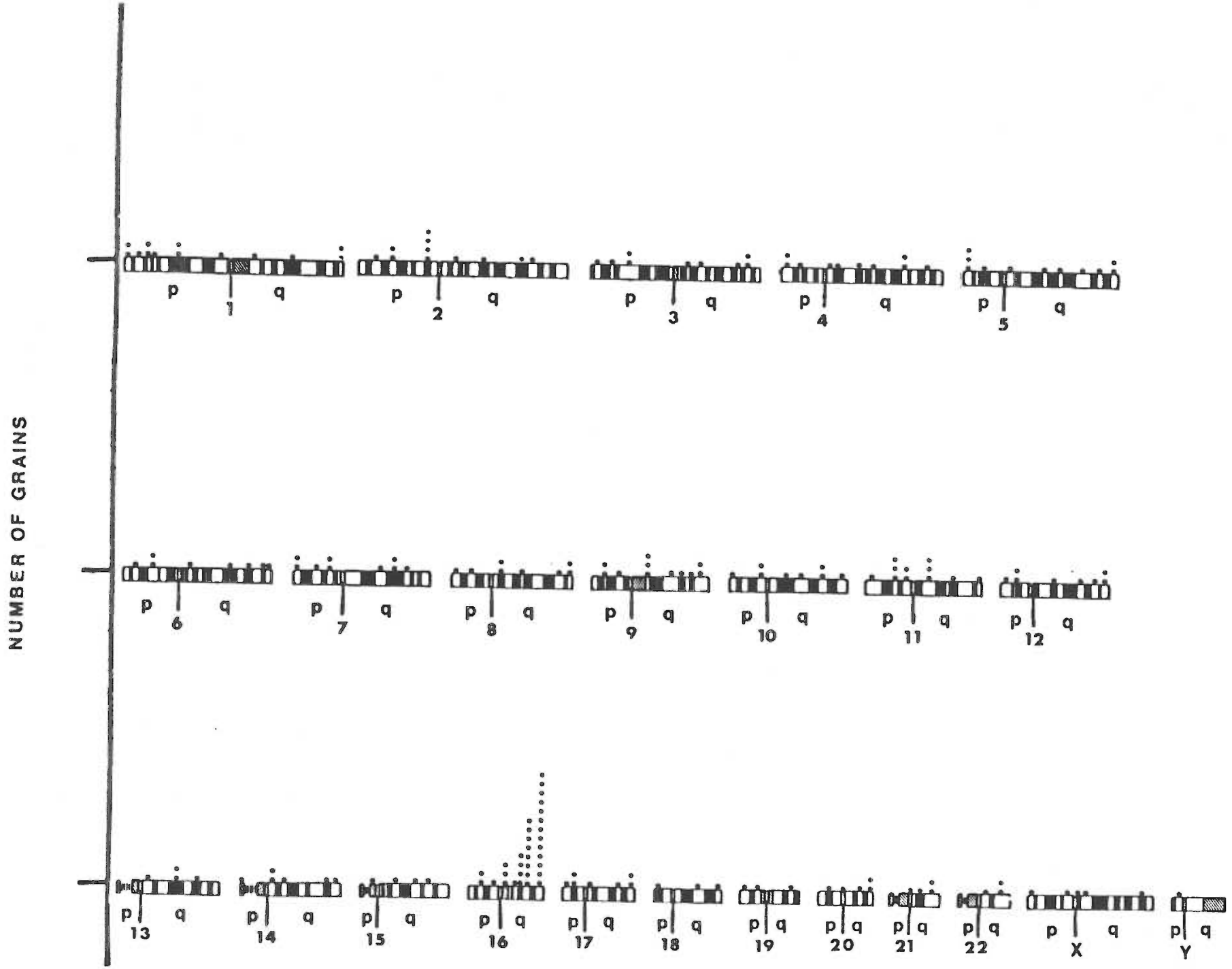


Figure 3. Histogram of chromosomal distribution of silver grains from *in situ* hybridization of phage Φ8-9 to chromosomes of a normal male. Each dot represents a grain observed over a specific chromosomal region. 21 of 110 cells examined had a grain localized to the region 16q22-24.


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

REFERENCES

Barker D, Holm T, White R (1984) A locus on 11p with multiple restriction length polymorphisms. Am J Hum Genet 365:1159-1171.

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

BMBiochemica (1985) Molecular grade intestinal alkaline phosphatase: the problem solver. BMBiochemica 2;(3):8.

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

Bufton L, Mohandas TK, Magenis RE, Sheehy R, Bestwick RK, Litt M (1986a) A highly polymorphic locus on chromosome 16q revealed by a probe from a chromosome specific cosmid library. Hum Genet 74:425-431.

Bufton L, Bruns GAP, Magenis RE, Tomar D, Shaw D, Brook D, Litt M (1986b) Four restriction fragment length polymorphisms revealed by probes from a single cosmid map to chromosome 19. Am J Hum Genet 38:447-460.

Dracopoli NC, Houghton AN, Old LJ (1985) Loss of polymorphic restriction fragments in malignant melanoma: implications for heterogeneity. PNAS 82:1470-1474.

Egeland JA, Gerhard DS, Pauls DL, Sussex JN, Kidd KK, Allen CR, Hostetter AM, Housman DE (1987) Bipolar affective disorders linked to DNA markers on chromosome 11. Nature 325:783-787.

Fearon ER, Feinberg AP, Hamilton SH, Vogelstein B (1985) Loss of genes on the short arm of chromosome 11 in bladder cancer. Nature 318:377-380.

Feinberg AP, Vogelstein B (1983) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal Biochem 132:6-13.

Fratini A, Simmers RN, Callen DF, Hyland VT, Tishfield JA, Stambrood PJ, Sutherland GR (1986) A new location for the human adenine phosphoribosyltransferase (APRT) gene distal to haptoglobin and Fra (16)(q23) (Fra16D) loci. Cytogenet Cell Genet 43:10-13

Friend SH, Bernards R, Rogel S, Weinberg R, Rapaport JM, Albert DM, Dryja TP (1986) A human DNA segment with properties of the gene that predisposes to retinoblastoma and osteosarcoma. Nature 323:643-646.

Frischauf AM, Lehrach H, Poutska A, Murray M (1983) Lambda replacement vectors carrying polylinker sequences. J Mol Biol 170:827-842

Gusella JF, Keys C, Varsanyi-Breiner A, Kao F-T, Jones C, Puck TT, Housman D (1980) Isolation and localization of DNA segments from specific human chromosomes. PNAS 77:2829-283.

Koufos A, Hansen MF, Copeland NG, Jenkins NA, Lampkin BC, Cavenee WK (1985) Loss of heterozygosity in three embroynal tumours suggests a common pathogenetic mechanism. Nature 316:330-334.

Lathrop GM, Lalouel J-M, Julier C, Ott J (1984) Strategies for multilocus linkage analysis in humans. PNAS 81:3443-3446

Litt M, Jorde LB (1986) Linkage disequilibria between pairs of loci within a highly polymorphic region of chromosome 2q. Am J Hum Genet 39:166-178.

Litt M, White R (1985) A highly polymorphic locus in human DNA revealed by cosmid derived probes. PNAS 82:6206-6210.

Maniatis T, Fritsch EF, Sambrook J (1982) "Molecular cloning: a laboratory manual". Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Marx JL (1985) Putting the human genome on the map. Science 229:150-151.

Monoco AP, Neve RL, Collett-Feener C, Bertelson CJ, Kurnit DM, Kunkel LM (1986) Isolation of candidate cDNAs for portions of the Duchenne Muscular Dystrophy gene. Nature 323:646-650.

Morton NE (1962) Segregation and Linkage, in "Methodology in Human Genetics", (WJ Burdette, Ed.) pp17-52, Holden Day, San Francisco.

Royer-Pokora B, Kunkel LM, Monaco AP, Goff SC, Newburger PE, Baehner RL, Sessions-Cole F, Curnutte JT, Orkin SH (1986) Cloning the gene for an inherited human disorder - chronic granulomatous disease - on the basis of its chromosomal location. Nature 322:32-38.

Sealey PG, Whittaker PA, Southern EM (1985) Removal of repeated sequences from hybridization probes. Nucleic Acids Res 13:1905-1922

Silhavy TJ, Berman ML, Enquist LW (1984) Experiments with gene fusions. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

St George-Hyslop PH, Tanzi RE, Polinsky RJ, Haines JL, Nee L, Watkins PC, Myers RH, Feldman RG, Pollen D, Drachamn D, Growdon J, Bruni A, Foncin J-F, Salmon D, Frommelt P, Amaducci L, Sorbi S, Piacentini S, Stewart GD, Hobbs WJ, Conneally PM, Gusella JF (1987) The genetic defect causing familial Alzheimer's disease maps on chromosome 21. Science 235:885-890.

Tanzi RE, Gusella JF, Watkins PC, Bruns GAP, St George-Hyslop P, Van Keuren ML, Patterson D, Pagan S, Kurnit DM, Neve RL (1987) Amyloid βprotein gene: cDNA, mRNA distribution, and genetic linkage near the Alzheimer locus. Science 235: 880-884.

Tusi LC, Buchwald M, Barker D, Braman JC, Knowlton RG, Schumm JW, Eiberg H, Mohr J, Kennedy D, Plavsic N, Zsiga M, Markiewicz D, Akots G, Brown V, Helms C, Gravius T, Parker C, Rediker K, Donis-Keller H (1985) Cystic fibrosis locus defined by genetically linked polymorphic DNA marker. Science 230:1054-1057.

Vogelstein B, Fearon ER, Hamilton SR, Feinberg AP (1985) Use of restriction fragment length polymorphisms to determine the clonal origin of human tumors. Science 227:642-645.

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

White R, Woodward S, Leppert M, O'Connell P, Hoff M, Herbst J, Lalouel J-M, Dean M, Vande Woude G (1985) A closely linked genetic marker for cystic fibrosis. Nature 313:382-384

Willard HF, Skolnick MH, Pearson PL, Mandel J-L (1985) Report of the Committee on Human Gene Mapping by Recombinant DNA Techniques. (8th International Workshop on Human Gene Mapping) Cytogenet Cell Genet 40:360-490.

III. MANUSCRIPT 2:

Six Polymorphic Loci Mapping to Human Chromosomal Region 11q21-qter

SUMMARY

Six polymorphic loci which map to human chromosomal region 11q22-qter are revealed by DNA probes isolated from a chromosome specific phage library. The library was constructed from a human x mouse somatic cell hybrid that has retained an 11q;16q translocation as the only human DNA. Phage with human DNA inserts were prehybridized with total human DNA prior to screening to identify probes which reveal RFLPs. Three probes each of which reveals a two allele polymorphism, two probes each of which detects two linked RFLPs, and one probe displaying closely linked restriction site and insertion/deletion RFLPs have been characterized. Family studies indicate that two of the probes are closely linked resulting in a highly polymorphic compound locus, with a population heterozygosity of 90%.

The six clones have been localized to specific chromosomal regions using a somatic cell hybrid mapping panel which divides 11q into seven discrete sections; *in situ* hybridization has been used to map one of the clones.

INTRODUCTION

Restriction fragment length polymorphisms (RFLPs) provide direct access to sequence variation in the human genome. Construction of detailed genetic linkage maps using RFLPs as marker loci is furnishing the tools necessary for the localization of genes through linkage analysis. Mapping of disease genes is important for several reasons. First of all, correlation of the segregation of a simple Mendelian trait with the inheritance of a specific chromosomal region defined by genetic markers holds the immediate promise of preclinical diagnosis through family studies. Polycystic kidney disease (1), cystic fibrosis (2-4) and Duchenne muscular dystrophy (5) are just a few of the single gene defects which can be diagnosed at the DNA level through the use of closely linked RFLPs. Localization of a defective gene to a well defined genomic region is a crucial first step towards the cloning and identification of that gene in the absence of other biochemical information. The recent cloning of sequences from the Duchenne muscular dystrophy locus (6), the chronic granulomatous disease gene (7), and the retinoblasoma/osteosarcoma locus (8) demonstrates the importance of mapping as a precursor to molecular identification of a gene.

Diseases with genetic components, but non-Mendelian inheritance patterns, have typically been refractory to linkage analysis. Genetic heterogeneity, multifactorial interactions and incomplete penetrance are all confounding factors which have impaired efforts to map the involved genes. Recently, the mapping of subsets of two multifactorial diseases by linkage to established marker loci emphasizes the importance of the development of a complete genetic linkage map. Detection of linkage between familial autosomal dominant Alzheimer's disease and markers on chromosome 21 has led to the discovery of a candidate gene for this disorder (9-10). In the second case, RFLP markers demonstrated genetic linkage between bipolar affective disorder in an Amish kindred and chromosome region 11p (11). This is the first molecular evidence for genetic involvement in a psychiatric disorder.

Primary linkage maps are now available for chromosomes X, 7, 12 and 13, and regions 6p and 11p (12). Some other genomic regions lag far behind in map construction. The long arms of chromosomes 11 and 16 are two regions particularly devoid of assigned polymorphic markers. As of the 8th International Human Gene Mapping Workshop (HGM8, 1985) chromosome region 16q had two assigned RFLPs. Since then we have mapped two highly polymorphic loci to that region (13,14). As of the same time, chromosome region 11q had just three polymorphic

loci assigned of which one, APOAI, has a polymorphism information content (PIC) value greater than 0.5 (15).

The advantages of mapping with highly polymorphic marker loci are well documented (16). PIC values greater than 0.5 are desirable, while RFLPs with PIC values approaching 1.0 provide ideal markers. The even spacing of highly polymorphic genetic markers at approximately 20 cM intervals would provide a complete genomic linkage map which would allow the subsequent placement of genes through linkage analysis (16). However, it is possible that not all regions of the genome have highly polymorphic individual loci. An effective alternative is to saturate unmapped regions with lower PIC RFLPs at a greater density. To a first order approximation, if a marker is informative half of the time, doubling the number of markers by using similar closely linked loci will be equivalent to using a single highly polymorphic marker, assuming no linkage disequilibrium (16).

Here we report the isolation and characterization of a series of six probes which reveal nine RFLPs mapping along the distal portion of 11q, from 11q22-11qter. One probe detects two RFLP loci, with a compound PIC value of 0.58. Another compound locus has a PIC value of 0.46. A third probe detects two loci in total linkage disequilibrium with a resulting PIC of 0.37. Each of the remaining three probes detect a two allele RFLP.

MATERIALS AND METHODS

Somatic Cell Hybrids

DNA used to construct the chromosome specific library was obtained from a previously described mouse x human somatic cell hybrid (CF52) that retained a single translocated human chromosome, t(11;16)(q13;p11), on a rodent background (13).

Cell hybrids for the genomic mapping panel were provided by Gail Bruns and have been previously described (17,18).

Cell hybrids for the chromosome 11 mapping panel were provided by Carol Jones (J1-11 and J1-44) and Tom Glaser (MC-1, TG 5D1-1 and G1-7). Cell lines J1-11 and J1-44 have been described elsewhere (19). MC-1 is a microcell hybrid made from GM3552, a human donor fibroblast with a reciprocal X-11 translocation (20) and CHTG49, a thioguanine-resistant Chinese hamster cell line (21). The MC-1 clone was selected in HAT + 1 M ouabain. It carries the 11-X translocation (11pter-q23::Xq26-qter) as the only human DNA. TG 5D1-1 is a Friend cell line derived from a hybrid (5D1) that carries intact human chromosomes 11 and X (22). It was back-selected against 6-thioguanine, consequently losing the X and a major portion of chromosome 11. It has been characterized with genetic markers by comparison to J1-44. G1-7 was also derived from 5D1. A retroviral marker with a neomycin

resistant gene (zipneo SV(X)1) was inserted into the long arm of the chromosome 11 carried by 5D1. Chromosomal fragments were then transferred to CHTG49 cells, and neomycin resistant clones were selected by exposure to G418. The resulting hybrid, G1-7, was characterized by hybridization with two genetic markers which map to 11q22.

Chromosome Specific Phage Library

The 11q;16q chromosome specific library was constructed and screened as described elsewhere (14).

Probe Preparation

All DNA probes were prepared by oligolabeling with [^{32}P] dCTP according to Feinberg and Vogelstein (23).

In the case of probes Φ2-14, Φ2-22, Φ2-25 and Φ6-3, whole phage DNAs are labeled and then prehybridized with total human DNA prior to use on Southern blots as previously described (18,24).

Probe Φ2-11-2.2 was prepared from phage Φ2-11 by Eco RI digestion of the phage DNA, followed by separation of the resulting fragments by agarose gel electrophoresis. The unique 2.2 kb restriction fragment was then cut out of the gel, oligolabeled directly in the agarose, and used as a probe without further purification.

Probe p2-7-1D6 was isolated as a subclone of phage Φ2-7. A Sau 3A digest of the phage DNA was ligated into

the Bam HI site of pSP65 (Promega Biotec; Madison, WI). Subclones of the phage were isolated on 2xYT/ampicillin plates (25) after transformation of the ligation mixture into the lambda resistant host strain, JPA101 (26). Single copy human subclones were identified by hybridization to insert DNA isolated from the parent phage, and by a lack of hybridization with a total human DNA probe. These subclones were then screened on Southern blots for their ability to detect the Taq I RFLP originally detected by Φ2-7.

Occasionally, some probes produce high background on Southern blots. In these instances, the signal to noise ratio is often improved significantly by first hybridizing the probe to a Southern blot of equal size, but of no experimental consequence (S. Kondoleon, pers. comm.). The probe is incubated in the mock hybridization for 30 minutes under the same conditions as the experimental hybridization, and then transferred to the Southern blot of interest for hybridization as usual.

Chromosomal Localization

Initial localization to chromosome 11 was achieved by use of a rodent x human somatic cell hybrid mapping panel, which was described previously (17,18). A more precise localization was obtained with a human x rodent somatic cell hybrid mapping panel. Hybridization to the panel allows assignment of the probes to one of seven

discrete regions of 11q by analysis for presence or absence of hybridization to fragments defined by various breakpoints along the chromosome.

Regional localization for one probe, Φ2-22, was verified by *in situ* hybridization to human metaphase chromosomes as described previously (13,18). The probe was prepared for *in situ* hybridization by an oligolabeling procedure described elsewhere (14).

Linkage Analyses

Lymphoblast cell lines from 9 three-generation Utah reference families, each of which includes all four grandparents and at least six grandchildren, were obtained from the Institute for Medical Research, Camden, NJ.

These data were used to determine linkage relationships among pairs of marker loci. Genetic linkage analyses were carried out using the LINKAGE computer program (27). LOD scores were calculated for recombination fractions (θ) of 0.0, 0.1, 0.2, 0.3, 0.4 and 0.5. When positive linkage scores are detected, the program reiterates through smaller increments of θ until the maximum LOD score is calculated.

In cases where more than one locus is detected by a single probe, the pairs were analyzed for linkage equilibrium parameters. Chi-square (X^2) values were calculated for goodness of fit for observed versus

expected haplotype frequencies. The associated probabilities, p, were calculated for one degree of freedom.

Polymorphism Information Content

PIC values were calculated according to Botstein *et al*, using allele frequencies for single diallelic loci, and haplotype frequencies for compound loci (28).

Results

Polymorphic Loci

Six recombinant phage which detect a total of nine polymorphic loci mapping to chromosome region 11q were isolated from a library specific for human chromosome arms 11q and 16q. Table 1 summarizes the data pertinent to the assignment of these polymorphisms to six loci.

Φ2-25 and p2-7-1D6 each detect single restriction site polymorphisms. In the case of Φ2-25, this can be determined because the probe overlaps the variable Msp I restriction site, and three variable bands are detected. The cosegregant 3.2 + 1.15 kb fragments are derived from cleavage of the 4.35 kb fragment when the restriction site is present (Fig.1).

The subclone p2-7-1D6 hybridizes to 6.4 and 4.25 kb fragments found at that restriction site polymorphism. The 2.15 kb fragment which also results from cleavage of

the 6.4 kb fragment was detected with a separate subclone of the parent phage, Φ2-7 (data not shown).

With Msp I digested DNAs Φ2-22 detects two loci, designated A and B. Locus A has two alleles, seen as 7.35 kb and 7.1 kb bands. Variable fragments with a 0.25 kb size difference are also seen with Taq I, Pst I and Bgl II, indicating that the RFLP is due to an insertion or deletion of a 0.25 kb fragment between the restriction sites (Fig.2). The B locus of Φ2-22 appears to be a closely linked restriction site polymorphism.

The diallelic loci detected by Φ2-11-2.2 ,Φ2-14 and Φ6-3 are most likely restriction site polymorphisms, as no evidence has been found to imply an insertion/deletion mechanism.

Three of the probes, Φ2-14, Φ2-22 and Φ6-3 each detect two diallelic loci. Methods for assignment of alleles to specific loci vary, depending on the probe. The two loci revealed by Φ2-14 are each detected with a different enzyme, eliminating ambiguity. In the case of Φ2-22, assignment of allelic fragments to specific loci was based on analysis of the same individuals with Msp I, which detects both loci, and Taq I which detects only locus A, an insertion/deletion polymorphism. With both enzymes, the fragment lengths vary by 0.25 kb and each individual displays the same genotype. For the probe, Φ6-3, a constant band obscured one of the allelic fragments when whole phage DNA was used as the probe

(Fig.3). Subdivision of the probe with Hind III separates the two loci. A unique 2.7 kb Hind III fragment detects the A locus of Φ6-3, while the B locus is detected by the 5.0 kb fragment (Fig.4). The remaining three probes, Φ2-11-2.2, Φ2-25 and p2-7-1D6 each detect a two allele RFLP. In all cases, allelism was confirmed by observing the segregation of the fragments in several Utah reference families. Codominant Mendelian inheritance was observed for each locus.

Designations from the Human Gene Mapping Library (15) for these six loci (including the three compound loci) are shown in Table 1.

Population Studies

Southern blot analyses of DNA from panels of unrelated Caucasians were done for each probe. The sample size, fragment patterns observed and their associated allele frequencies are included in Table 1. In each case, observed and expected genotype frequencies are in good agreement, indicating that each locus is in Hardy-Weinberg equilibrium.

In the cases where single probes detect two loci, haplotype frequencies were calculated based on haplotypes inferred from the genotypes of unrelated individuals. Analysis of observed versus expected haplotype frequencies for linked loci indicates that the incidence of linkage disequilibrium is highly variable (Table 2).

There is good agreement between observed and expected haplotype frequencies for D11S36, the locus detected by Φ2-14, indicating no significant linkage disequilibrium. Haplotype analysis for the locus D11S35, revealed by Φ2-22, shows strong disagreement between observed and expected frequencies, with the A1B2 and A2B1 haplotypes being underrepresented. The loci of D11S85 were found to be in total linkage disequilibrium, with a complete absence of the A1B2 and A2B1 haplotypes.

The relative effects of linkage disequilibrium on population heterozygosity are reflected in the PIC values calculated for the compound loci revealed by Φ2-14, Φ2-22, and Φ6-3 (Table 3). PIC values were also calculated for the individual loci, revealing how linkage disequilibrium affects the enhancement of PIC to varying extents when compound loci are considered.

PIC values for single loci not closely linked, are also included in Table 3.

Linkage Analyses

Detection of two or more loci by a single recombinant probe suggests, but does not prove, close linkage of those loci. Southern blot analyses of the nine Utah reference families confirmed close genetic linkage of loci detected by each of the three single probes. Maximum LOD scores resulting from these analyses are presented in the first three rows of Table 4. In

each case, the LOD score is consistent with close linkage.

In addition there is physical evidence of colinearity of the cloned insert of Φ2-14 and the genome. The probe insert is shown to be colinear with a single segment of the human genome by its hybridization to a discrete number of nonpolymorphic restriction fragments equal to the number of insert-derived fragments generated by that restriction enzyme. To demonstrate colinearity, the fragments detected by genomic Southern blot analysis must be of equal or larger size than the insert fragments. Eco RI digestion of Φ2-14 produces insert-derived fragments of 3.7, 1.9, 1.45 and 1.35 kb. A genomic Southern blot prepared with Eco RI and probed with Φ2-14 detected fragments of 5.0, 4.0, 1.45 and 1.35 kb, indicating that the insert was derived from a single segment of the genome. Loci internal to that segment are therefore physically and genetically linked.

Data from the nine Utah reference families were used to determine the extent and range of genetic linkage between each of the possible pairs of loci. Maximum positive LOD scores, and the associated values for θ are included in Table 4. Linkage is not indicated for the other pairwise combinations, with negative LOD scores generated for all values of θ analyzed.

Chromosomal Localization

Initially, all six probes were mapped to chromosome 11 by hybridization to Southern blots containing Hind III digested DNAs from 25 cell lines of a somatic cell hybrid mapping panel. Each probe hybridized to the eight cell lines which contain chromosome 11, and were discordant for all other chromosomes, as illustrated in Fig. 5, bottom panel, and Table 5.

A more precise localization was obtained by hybridization to a somatic cell hybrid mapping panel blot that contains Taq I digested DNAs from five cell lines, each containing different regions of chromosome 11 (Fig. 5 top panel). The mapping panel effectively divides 11q into seven regions. The results of the hybridizations for the six probes are shown in Table 6 and Figure 6.

Probe Φ6-3 has two possible regions of localization. The probe hybridizes to MC-1, but it does not hybridize to G1-7, indicating that it must be proximal to 11q23 and distal to 11q22, excluding the portion of 11q22 defined by G1-7. Therefore, it resides in that region, but it is either proximal or distal to G1-7.

Φ2-22 shows the same mapping panel hybridization pattern as Φ6-3, but is thought to reside proximal to the region defined by G1-7, based on *in situ* hybridization data. A representative, sequentially stained metaphase spread from the *in situ* hybridization study is shown in

Fig. 7. The silver grains revealing hybridization of the probe are observed with standard stain while the chromosomes to which the probe has annealed are identified by fluorescent R-banding. Fig. 8 summarizes the results of scoring 127 labeled metaphase spreads from a normal male. 18 of the 127 metaphases (14%) showed hybridization to 11q21-22; the remainder of the label was distributed randomly over the chromosomes.

DISCUSSION

Six phage clones which detect RFLPs mapping to the distal portion of chromosome region 11q have been isolated from an 11q;16q chromosome specific library. Three of the probes, Φ2-14, Φ2-22 and Φ6-3 each detect two closely linked diallelic loci, which were analyzed as compound loci. For Φ2-14 and Φ2-22, the creation of compound loci significantly increases the value of the probe as indicated by higher PIC values for the compound loci versus the PIC of a single diallelic locus. However, significant linkage disequilibrium between the A and B loci of Φ2-22 results in a lower compound PIC value than would otherwise be expected. The observed frequencies for the A1B1 and the A2B2 haplotypes are considerably higher than predicted from the allele frequencies, resulting in a larger than expected number of doubly homozygous individuals, and an

underrepresentation of heterozygotes for the compound locus.

The two loci detected by Φ6-3 are in total linkage disequilibrium. As a result, there is no benefit in treating the loci as a compound locus. In fact, since there is a 2.5 kb constant band obscuring an allele of locus B when whole phage DNA is used as a probe, it is best to use only locus A, alleviating any need to subdivide the probe prior to use.

Linkage analysis between pairs of loci provides preliminary evidence of genetic linkage for several of the markers. There is evidence that the compound loci D11S35 and D11S36 are closely linked and can be further compounded to substantially increase the information content of that locus. If treated as a single locus, the population heterozygosity is 90%.

Positive LOD scores at larger recombination fractions were also generated for some of the other pairs of loci. There is significant evidence of linkage between D11S84 and D11S35 at a recombination fraction of 0.15. There is also a suggestion of linkage between D11S84 and D11S36 at a recombination fraction of 0.10. It is clear that additional informative meioses are needed to estimate these genetic distances accurately. However, these genetic data are consistent with the physical data resulting from the mapping panel localizations. Also, more family studies will resolve

the problem of localizing D11S85 to the region either proximal or distal to the segment designated by the cell line G1-7. Using multipoint linkage analysis, it should also be possible to determine the genetic order of these loci. This will provide the basis for a genetic linkage map for the distal portion of 11q.

The remaining LOD scores were either negative or weakly positive. The positive LOD scores were seen for pairwise analyses between D11S83 and each of the loci D11S35, D11S36, and D11S84. Additional informative meioses will be necessary before any conclusions regarding linkage can be made for these pairs of loci.

The regional assignments of the probes described here indicate that they all derive from approximately the distal one-half of the long arm of chromosome 11, based on cytogenetic estimates of the physical sizes of these chromosome segments. Although the precise locations must be verified, it is surprising that six clones selected at random from a chromosome specific library all map to the distal half of a chromosome arm. A similar situation was recently reported by Gilliam *et al.*, for chromosome region 4p (29). Forty random polymorphic markers detected by single copy probes isolated from a flow-sorted chromosome 4 specific library were found to be unequally distributed along the short arm of chromosome 4. Sixty-eight percent of the polymorphic clones map to the region 4pter-p15.1; approximately the distal one-half

of the short arm. Although the authors can only speculate as to the significance of this unexpected distribution, it bears an interesting resemblance to the situation seen thus far on 11q.

In addition to the probes discussed here, two other polymorphic clones have been mapped from the chromosome specific library reported here. One clone is localized to the distal half of 11q (11q22) (J. Kellogg, pers. comm.), and the other to chromosome 16 (14). Since cytogenetic estimates also indicate that approximately one-half of the human genetic material in this library derives from chromosome region 16q (13,14), these results may suggest an underrepresentation of RFLPs on 16q, as well as a nonrandom distribution of RFLPs along 11q.

There are several possible explanations for these observations. The more interesting possibility is that our results reflect a nonrandom distribution of RFLPs on the long arms of chromosomes 11 and 16. There is also the possibility that the eight clones isolated so far are not reflective of the library content by coincidence, or because of an unexpected bias in the selection process. Other possible explanations include a nonrandom distribution of Sau 3a fragments along these chromosome arms, or that the library does not contain equal representation of all portions of the chromosome arms.

If the distribution of RFLPs along 11q and 16q are actually skewed, then probes from the library which

detect nonpolymorphic regions should be found to map proximal to the RFLPs described here or to 16q. Future isolation of polymorphic clones from this library will include mapping of both polymorphic and nonpolymorphic human loci on 11q in an effort to determine library constitution and to help assess the significance of the distribution of the RFLP markers described here.

The probes described here will be used to type the CEPH families (30), facilitating construction of a linkage map of 11q. They will also be available to researchers on a collaborative basis.

Acknowledgments

Excellent technical assistance was provided by Ruth Litt, Elaine Ostrander, Miyoung Kwak and Gwen Hansen. We thank Gail Bruns for providing the somatic cell hybrid mapping panel, and Christopher Dubay for computer analysis of LOD scores. We also thank Sophia Kondoleon and Elaine Ostrander for critical comments on the manuscript. This work was supported by NIH Grant RO1-GM32500, Maternal and Child Health Grant MCT920, and the Retinitis Pigmentosum Foundation Fighting Blindness. Cheryl Maslen was supported in part by a Research Fellowship from the Oregon Medical Research Foundation and the N.L. Tartar Trust Fund.

Table 1

Probe	Locus	Enzyme	Constant Bands (kb)	Alleles	Polymorphic Bands (kb)	Allele Frequencies
p2-7-1D6	D11S84	Taq I	-	A1	6.4	0.28 (74)
				A2	4.25	0.72
ϕ2-11-2.2	D11S34	Msp I	-	A1	8.9	0.64 (50)
				A2	7.6	0.36
ϕ2-14	D11S36	Rsa I	1.45	A1	2.95	0.59 (74)
				A2	2.1	0.46
		Pvu II	7.9	B1	12.5	0.89 (46)
				B2	11.5	0.11
ϕ2-22	D11S35	Msp I	6.7	A1	7.35	0.54 (56)
				A2	7.1	0.46
				B1	6.5	0.49
				B2	6.35	0.51
ϕ2-25	D11S83	Msp I	9.3, 5.0, 3.2, 3.0, 1.25, 1.12	A1	4.35	0.32 (60)
				A2	3.2 + 1.15	0.68
ϕ6-3	D11S85	Msp I	2.5, 1.7	A1	9.5	0.43 (70)
				A2	3.85	0.57
				B1	2.75	0.43
				B2	2.5	0.57

The numbers in parentheses are the number of chromosomes in the sample.

Table 2

EQUILIBRIUM ANALYSES BETWEEN LINKED LOCI

Locus	X^2	N[a]	p[b]
D11S36	2.1	30	0.16
D11S35	23.4	56	0.0001
D11S85	42.1	42	<<0.0001

[a]N is the number of chromosomes analyzed.

[b]p is the probability calculated for 1 degree of freedom.

Table 3

POLYMORPHISM INFORMATION CONTENTS

Locus	Insert Size (kb)	PIC/Locus A	PIC/Locus B	Compound PIC
D11S36	8.4	0.370	0.177	0.460
D11S35	15.0	0.373	0.375	0.580
D11S85	10.3	0.370	0.370	0.370
D11S83	16.5	0.375	-	-
D11S34	11.0	0.350	-	-
D11S84	0.93	0.322	-	-

Table 4

LINKAGE ANALYSES

Linkage Group Loci	Number of Informative Meioses	Recombination Fraction	Maximum LOD Score
D11S35 A/B	24	0.00	7.22
D11S36 A/B	9	0.00	2.97
D11S85 A/B	12	0.00	3.61
D11S35/D11S36	18	0.00	1.80
D11S84/D11S36	13	0.10	1.46
D11S84/D11S35	29	0.15	2.76
D11S36/D11S83	13	0.20	0.55
D11S35/D11S83	19	0.35	0.21
D11S84/D11S83	32	0.45	0.09

Table 5. Chromosome contents of hybrid cell lines (Fig.5a). 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 or by a cytogenetic abnormality; (±), chromosome is present in less than 15% of cells and/or the isozyme or DNA probe characteristic of the chromosome is 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 was excluded for analysis.

The column designated M indicates the presence or absence of human-specific Hind III fragments on Southern blots probed with pDP2-25 .

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

Lane 12 is the mouse parent cell line, and lane 25 is the hamster parent cell line.

Table 5

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	Scoring X	Y	M
1 G35F3	-	+	+	-	+	-	+	+	-	+	+	+	-	-	-	+	-	+	+	+	-	-	a	-	+
2 G35A3	-	p	-	±	-	-	+	+	+	+	+	-	-	±	-	+	+	-	+	+	+	+	a	-	+
3 G35F1	-	-	-	+	+	-	-	+	-	-	+	p	-	+	-	-	-	-	+	+	+	+	a	-	+
4 RRP3-6	+	+	+	+	+	+	+	-	-	+	+	+	-	-	+	+	+	+	R	-	+	-	+	+	+
5 RRP5-7	+	-	-	+	+	-	+	-	-	+	-	+	ND	+	-	-	+	+	-	+	-	-	+	R	-
6 G95A4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	±	-	-	+	-	-
7 G89E5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-
8 G24B5	+	-	+	+	-	-	+	-	-	+	+	+	+	+	-	+	-	-	+	-	+	+	+	-	+
9 G24A4	±	p	-	-	+	+	+	-	±	-	±	-	+	+	+	+	-	-	+	-	+	-	q	-	+
10 G24A9	-	+	-	+	-	+	-	+	+	+	-	-	±	+	+	-	±	+	+	-	+	-	q	-	-
11 G175	±	-	-	+	-	+	-	+	-	+	-	-	+	+	+	-	-	-	+	-	-	+	a	-	-
12 RAG	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
13 G35E4	-	-	-	R	-	-	-	-	R	-	-	-	-	-	+	-	-	-	+	-	-	-	a	-	-
14 G35B4	+	-	-	-	+	-	-	-	-	+	+	+	-	-	-	-	-	+	+	-	-	+	+	-	+
15 G35A4	+	-	+	+	-	+	-	+	+	-	-	-	-	+	-	+	-	-	+	+	+	-	a	-	-
16 G35C5	-	-	p	-	+	-	+	+	+	+	-	p	-	-	+	+	+	-	+	+	+	+	a	-	-
17 G35F5	±	-	+	+	-	+	-	-	+	-	+	-	+	+	-	+	+	+	+	±	+	+	a	-	+
18 G35B5	-	-	-	-	-	R	+	-	-	-	+	p	+	-	-	-	+	-	+	+	-	-	a	-	+
19 G35C4	-	-	-	-	-	+	-	-	R	-	-	+	-	+	-	-	-	±	+	+	+	-	a	-	-
20 G35C2	-	-	-	+	-	-	-	+	-	-	R	-	-	+	-	+	-	-	R	-	-	-	R	-	-
21 G35A2	-	-	+	±	-	+	-	-	-	-	+	-	-	+	-	-	-	-	+	+	+	+	+	-	+
22 G35E3	-	-	-	-	+	+	-	+	-	-	+	+	+	+	-	+	-	+	+	+	+	+	a	-	+
23 G35D2	±	-	-	-	+	+	-	-	+	-	+	+	-	-	+	-	-	+	R	+	-	+	a	-	+
24 G35D5	+	+	+	-	-	+	+	-	-	+	-	-	+	+	+	+	+	+	R	+	-	+	a	-	-
25 E36	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Discordancy Fractions:

Chromosomes

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	X	Y
6/19	9/21	7/22	9/20	6/23	8/22	7/23	15/23	9/20	12/22	0/21	5/20	7/21	11/22	12/23	8/23	7/22	8/22	8/20	8/21	6/23	8/23	3/8	11/22

Table 6

Probe Hybridization Patterns for 11q Mapping Panel

Probe	Hybrid Cell Lines MC-1	J1-11	J1-44	TG5D1-1	G1-7	Chromosomal Location	Comments
Φ2-22	+	-	+	+	-	11q22	see Figs.7&8
Φ2-14	+	-	+	+	+	11q22	distal to 2-22
p2-7-1D6	+	-	+	+	+	11q22	distal to 2-14?
Φ6-3	+	-	+	+	-	11q22-23	excludes G1-7
Φ2-25	-	-	+	+	-	11q23-qter	
Φ2-11-2.2	-	-	+	+	-	11q23-qter	

Figure 1. Southern blot of Msp I digested DNA samples from unrelated individuals probed with phage Φ2-25.

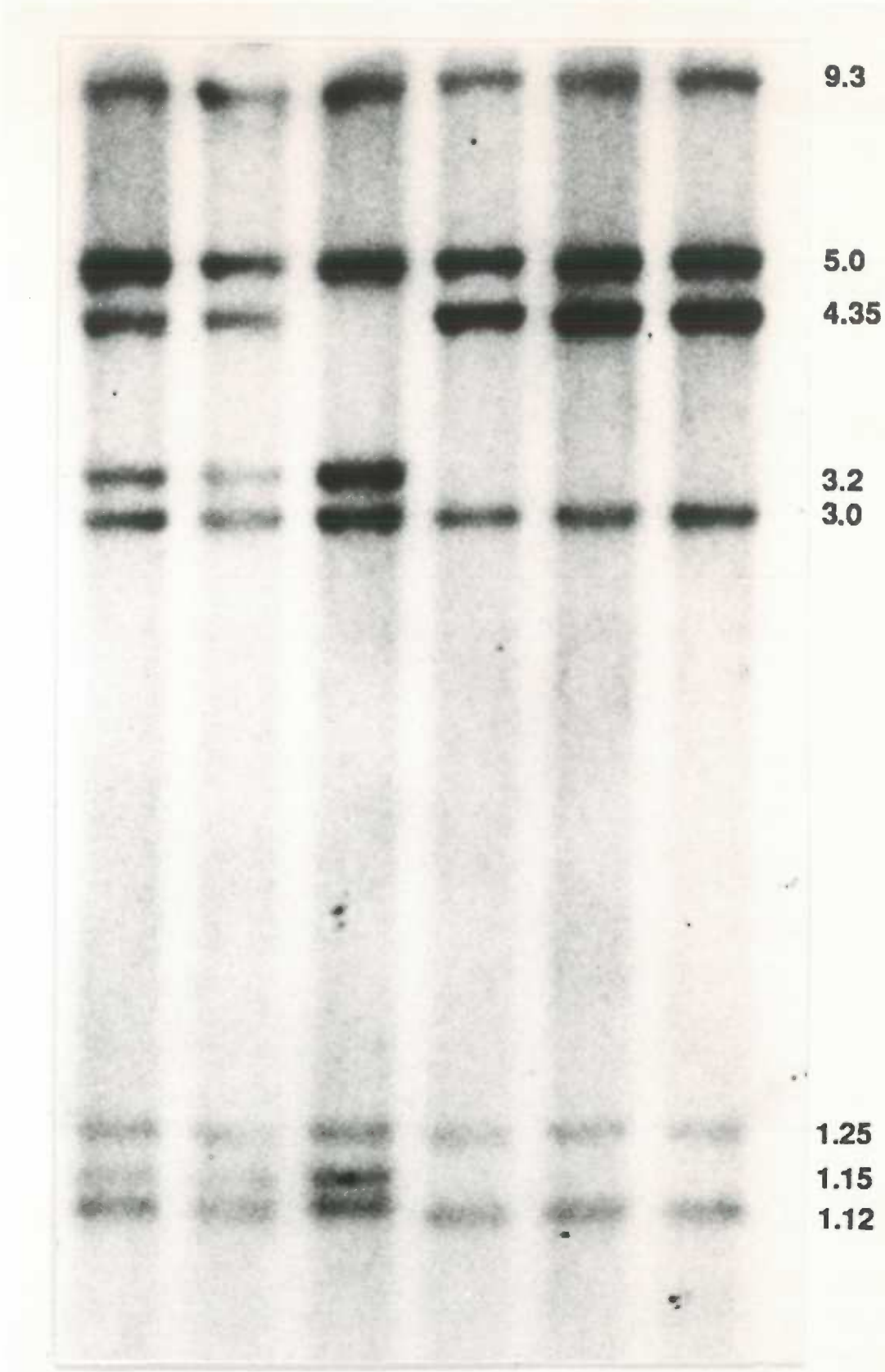

9.3
5.0
4.35
3.2
3.0
1.25
1.15
1.12

Figure 2. The top panel shows a Southern blot of Msp I digested DNAs from unrelated individuals probed with phage Φ2-22, revealing the closely linked A and B loci.

The bottom panel shows a Southern blot of Pst I digested DNAs from unrelated individuals also probed with phage Φ2-22, revealing locus A.
The 0.25 kb size difference between allelic fragments is seen at the A locus with both enzymes, due to an insertion/deletion mechanism.

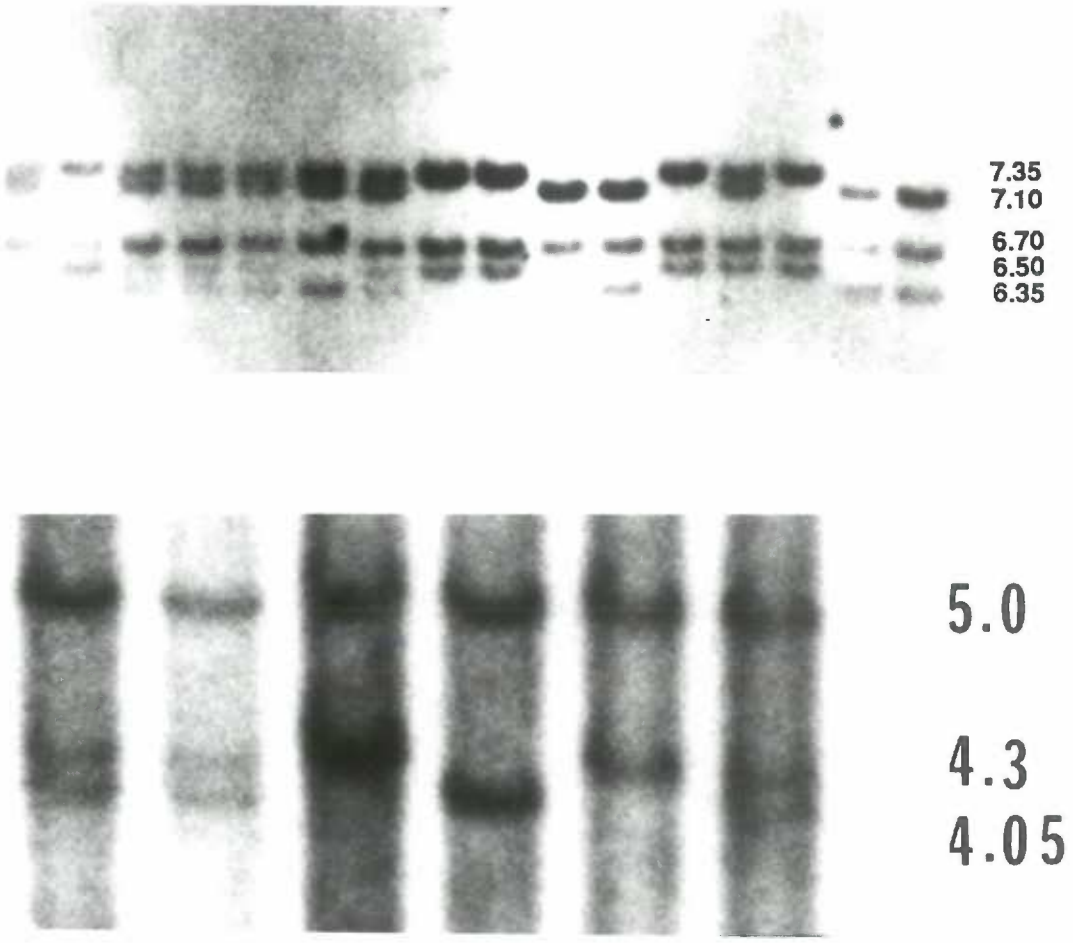

7.35
7.10
6.70
6.50
6.35
5.0
4.3
4.05

Figure 3. A Southern blot of Msp I digested DNAs from unrelated individuals probed with phageΦ6-3. A 2.5 kb constant band obscured the 2.5 kb allele of locus B.

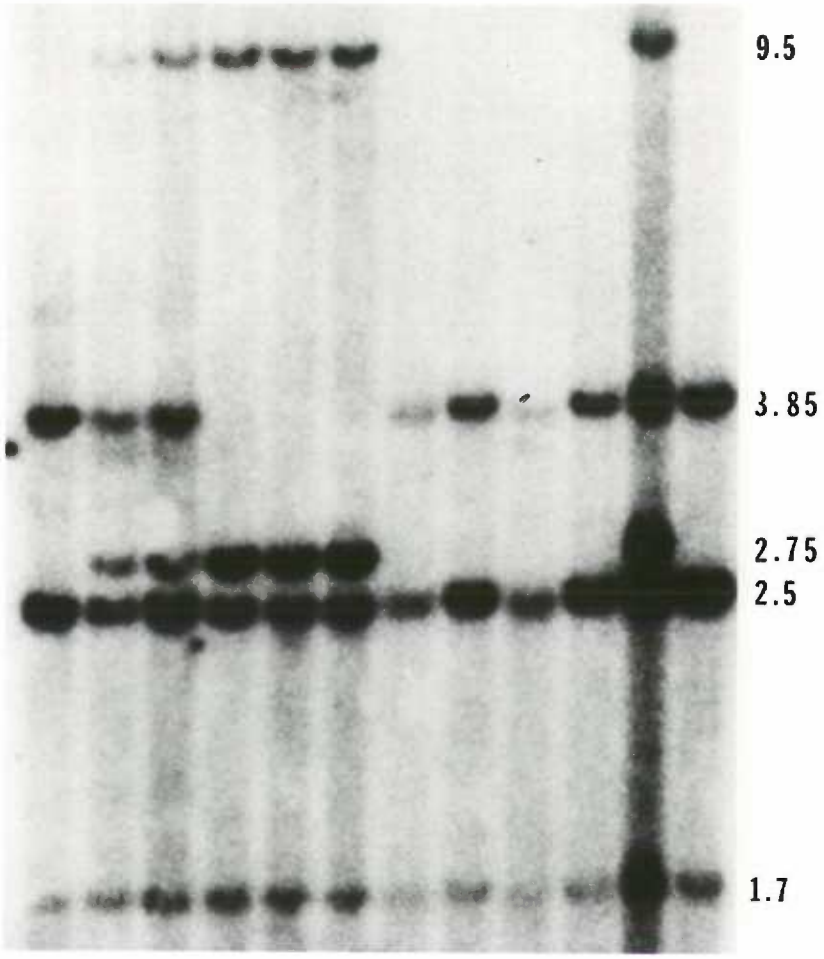
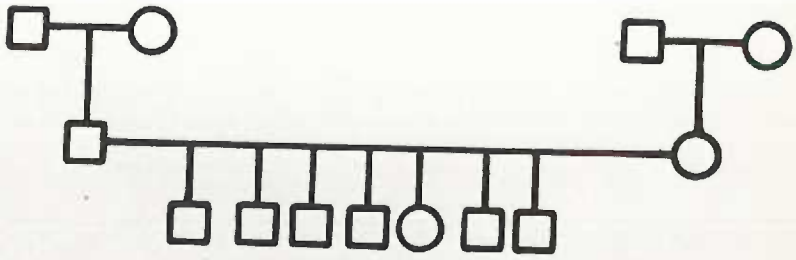

9.5
3.85
2.75
2.5
1.7

Figure 4. A Southern blot of Msp I digested DNAs from Utah reference family K1345 probed with the unique 2.7 and 5.0 kb Hind III restriction fragments of Φ6-3. The 2.7 kb probe reveals the 9.5 and 3.85 kb bands (locus A), and the 5.0 kb probe detects the 2.75 and 2.5 kb variable bands (locus B). Note the disappearance of the 2.5 kb constant band, allowing detection of the 2.5 kb variable band. A faint trace of the 2.5 kb constant band can be seen in some lanes due to a slight contamination of the probe with the 4.1 kb Hind III restriction fragment which detects that locus.



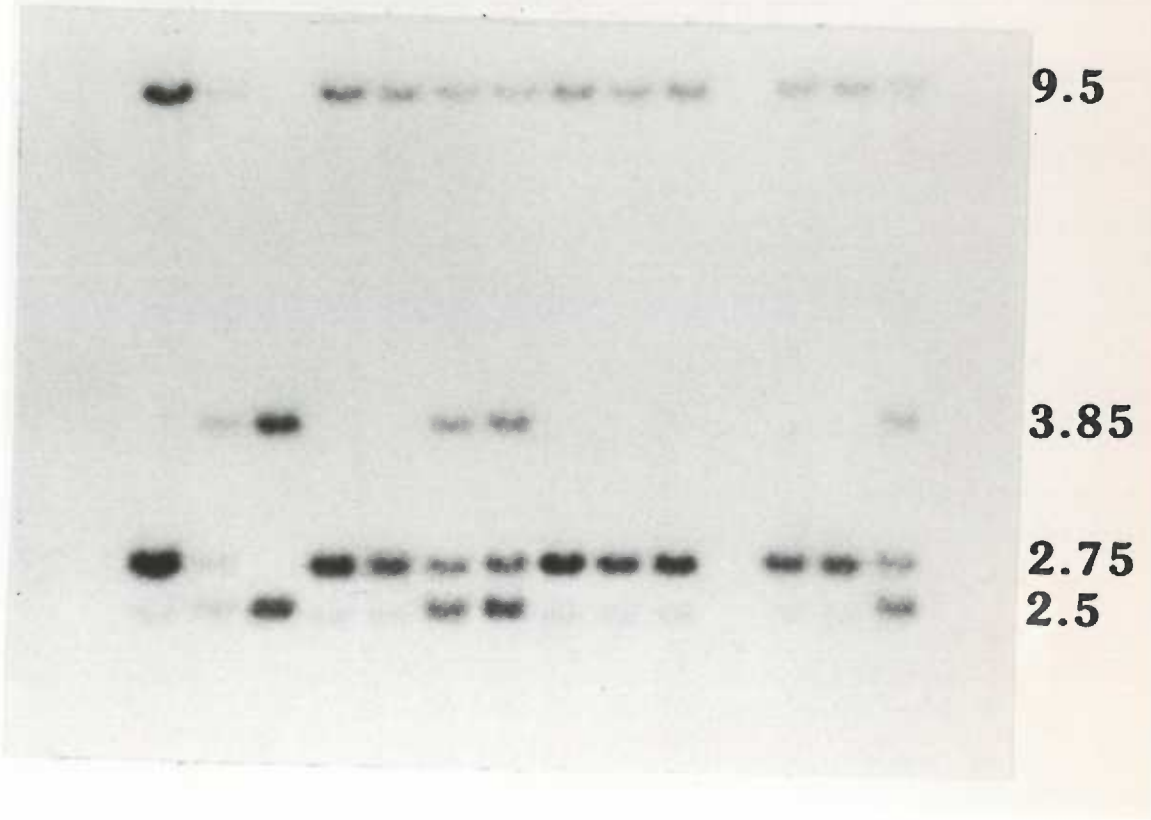

9.5
3.85
2.75
2.5

Figure 5. Top Panel: Southern blot of Taq I digested somatic cell hybrid DNAs probed with phageΦ2-22. Each lane contains DNA from a cell line with a different derivative chromosome 11. The chromosome complements of the cell lines are shown in Fig.6. The lane marked ss contains size standards. The results of the hybridization are tabulated in Table 6.

Bottom Panel: Southern blot of Hind III digested somatic cell hybrid DNAs probed with phageΦ2-25. The chromosome complements of the cell lines are shown in Table 5. The lanes marked ss contain size standards.

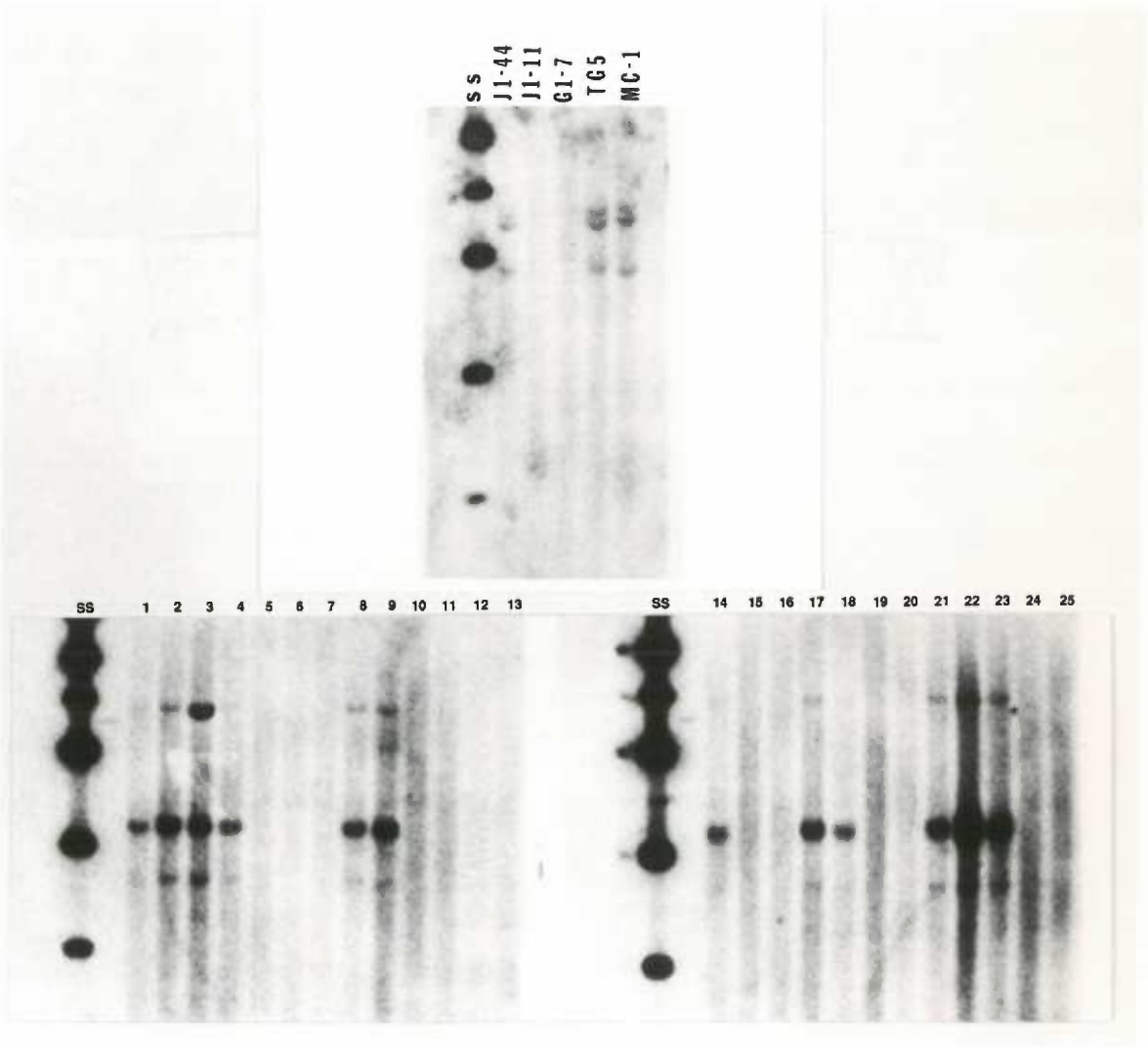

SS
J1-44
J1-11
G1-7
TG5
MC-1
SS
1 2 3 4 5 6 7 8 9 10 11 12 13
SS
14 15 16 17 18 19 20 21 22 23 24 25

Figure 6. An ideogram of chromosome 11 with a graphic representation of the human complement of each somatic cell hybrid used in the chromosome 11 mapping panel. The approximate breakpoints are indicated for each deletion. Estimated chromosomal locations for each locus as indicated by hybridization of the corresponding probe to the mapping panel are also shown (also see Fig. 5b and Table 6). D11S35, identified by probeΦ2-22, is positioned with regard to *in situ* hybridization data (Figs 7 & 8) as well as mapping panel information. D11S85, identified by probeΦ6-3, is shown at the two possible map placements.

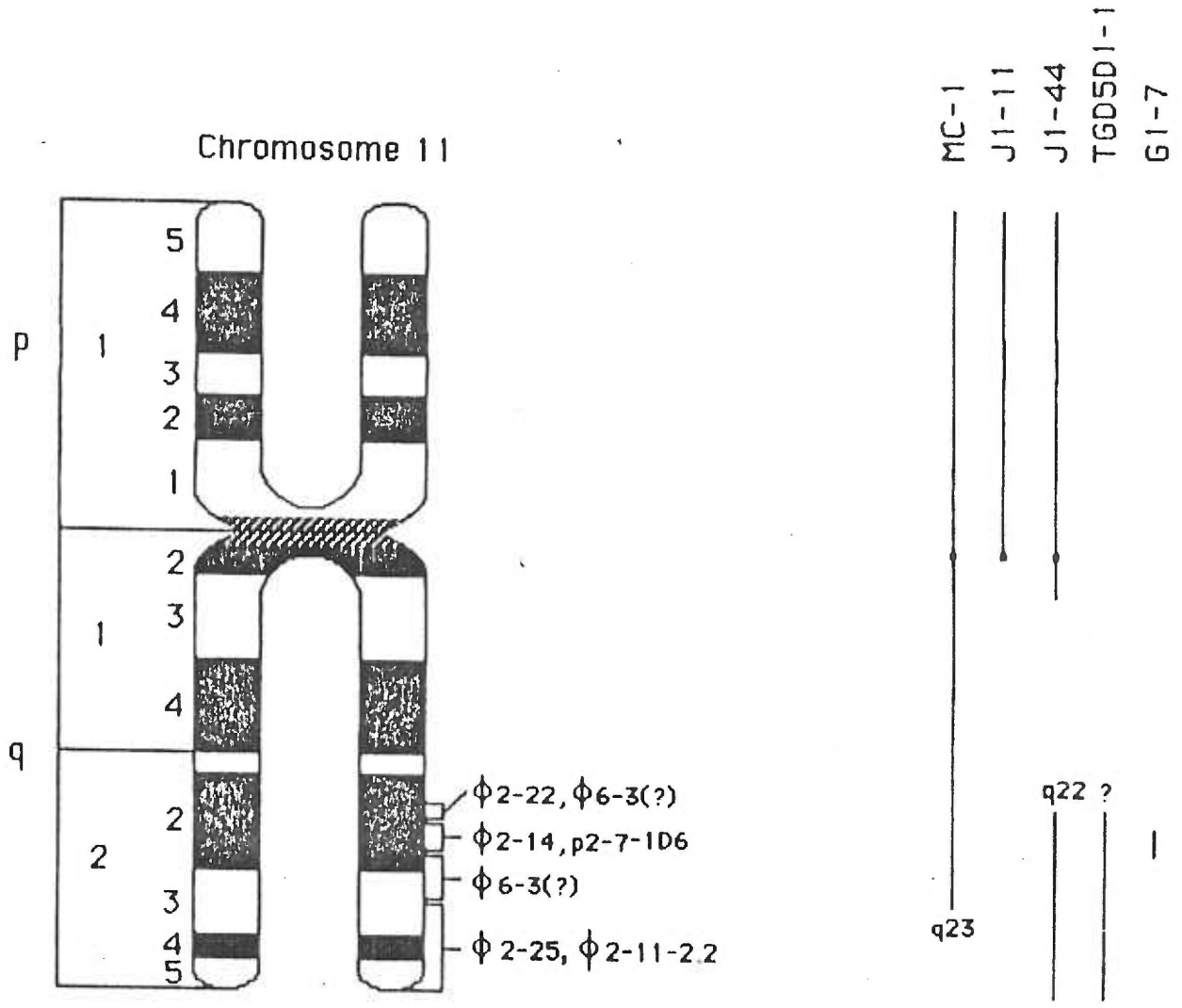
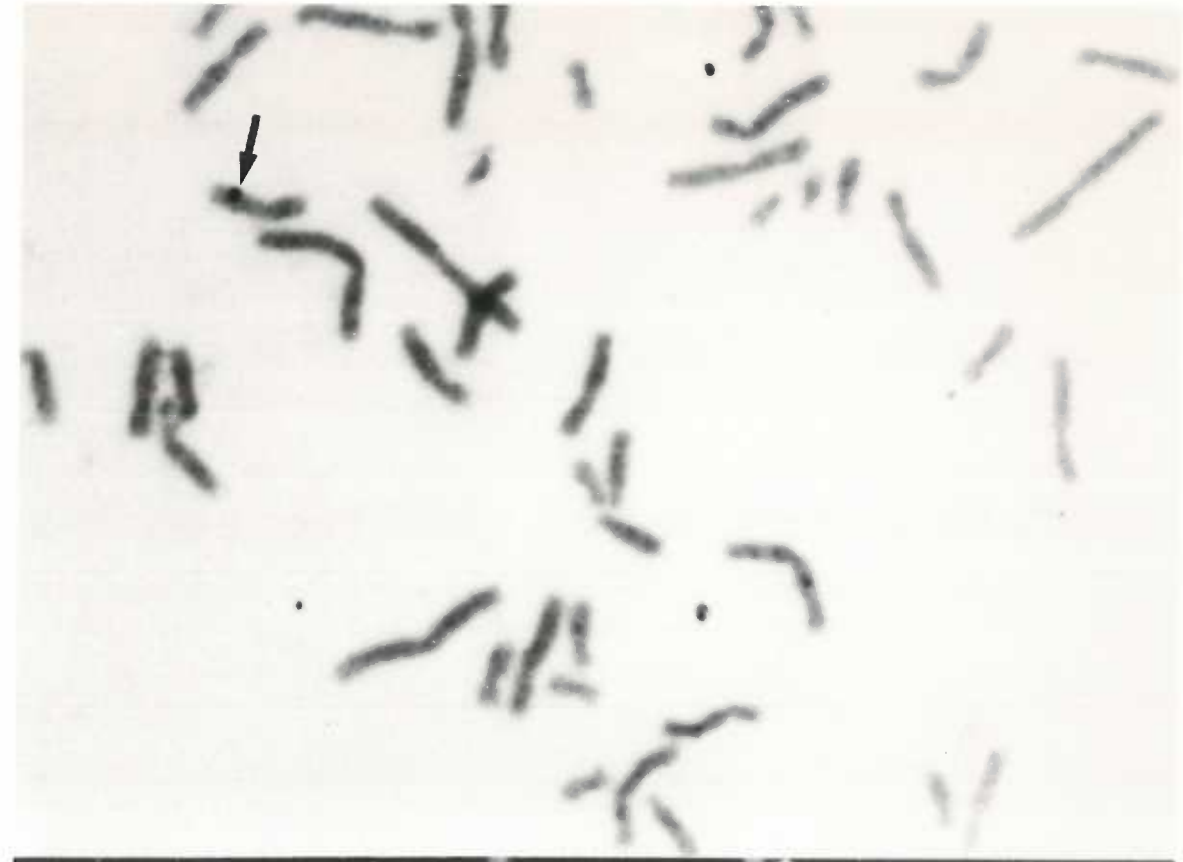

Chromosome 11
p
q
1
2
1
2
5
4
3
2
1
2
3
4
2
3
4
5
ϕ2-22, ϕ6-3(?)
ϕ2-14, p2-7-1D6
ϕ6-3(?)
ϕ 2-25, ϕ2-11-2.2
MC-1
J1-11
J1-44
TGD5D1-1
G1-7
q23
q22 ?

Figure 7. *In situ* hybridization of Φ2-22. Bottom panel: representative metaphase that has been R-banded following hybridization. Top panel: the same metaphase, destained and restained with Wright's stain to show silver grains. Arrow indicates silver grain located over the long arm of chromosome 11 and the corresponding location in the R-banded metaphase.



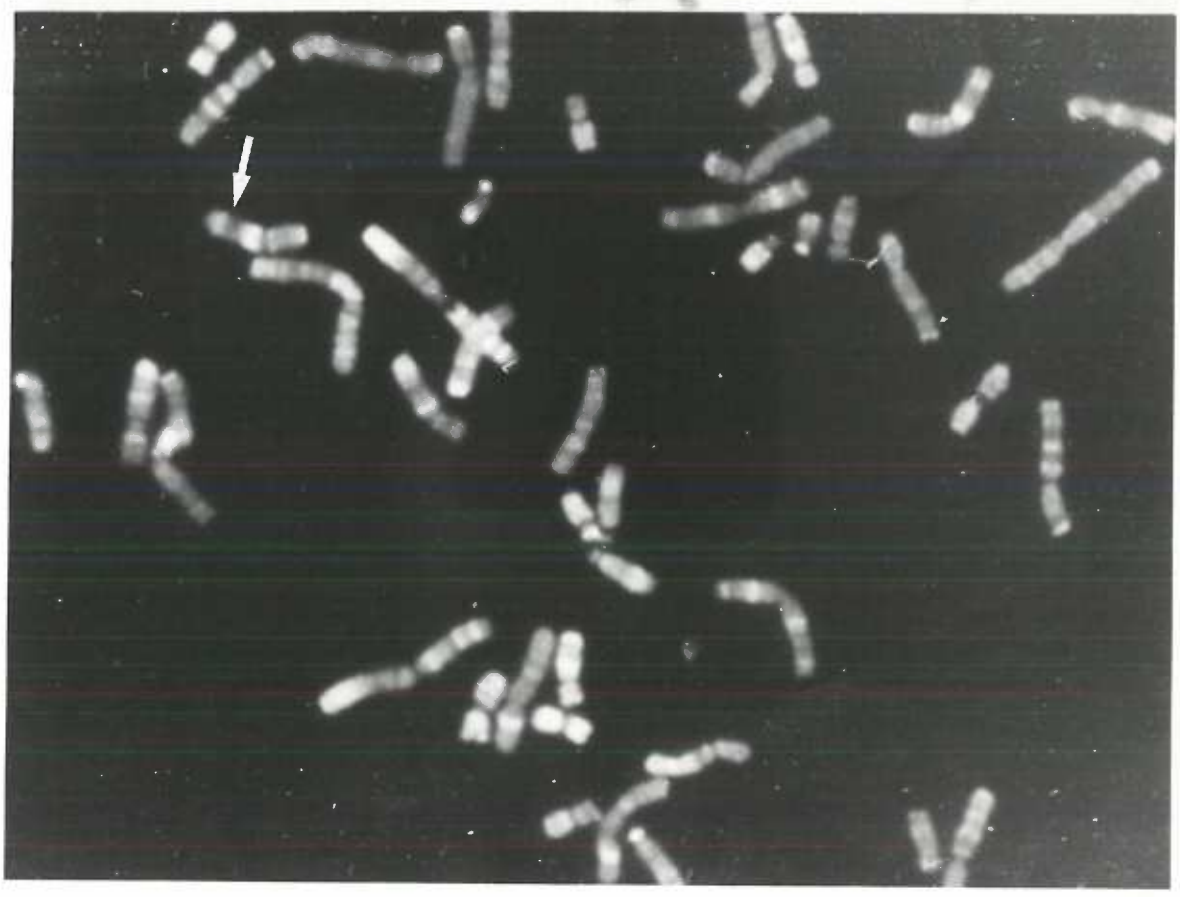
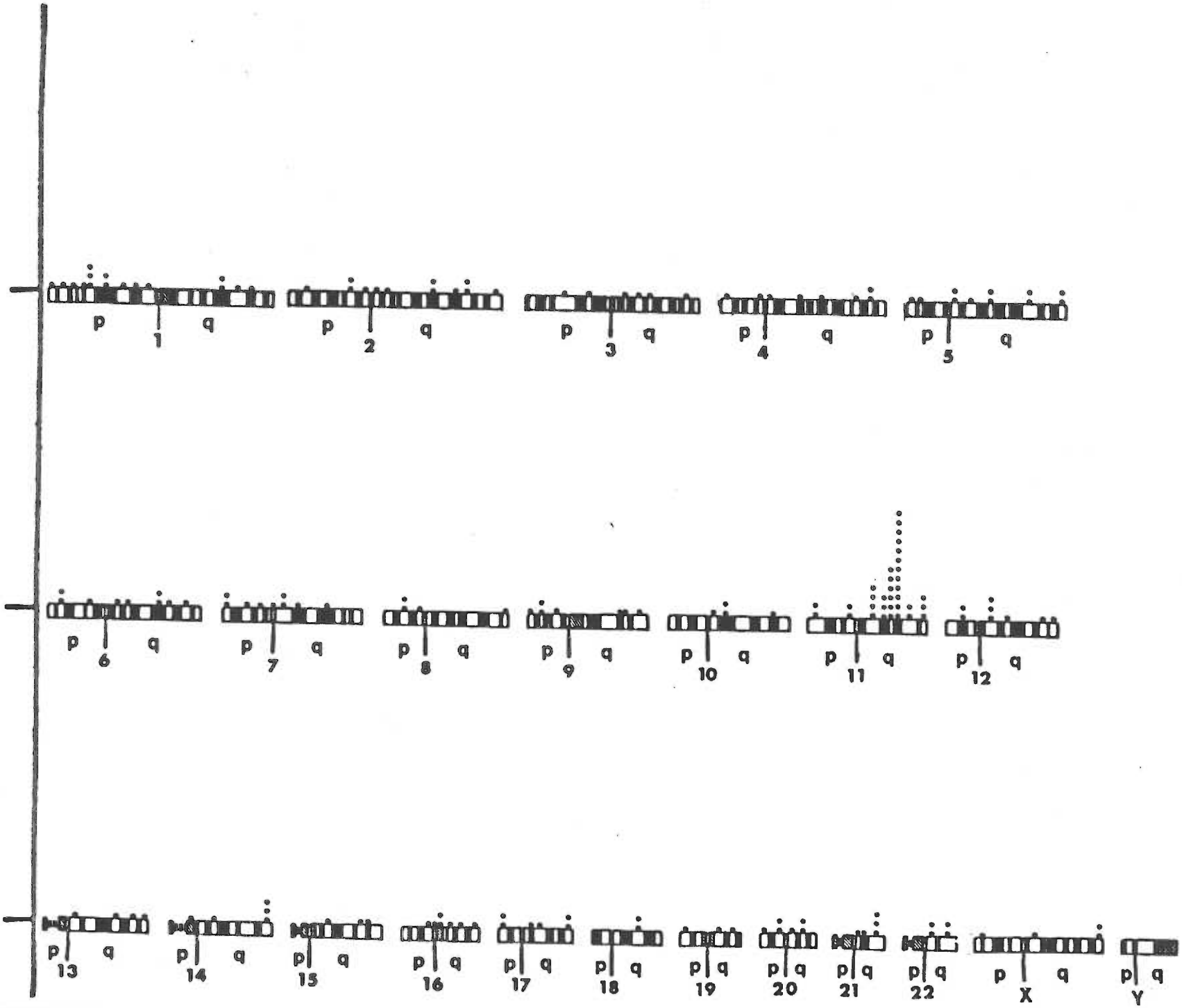


Figure 8. Histogram of the chromosomal distribution of silver grains from *in situ* hybridization of phage Φ2-22 to chromosomes of a normal male. Each dot represents a grain observed over a specific chromosomal region. 18 of 127 cells examined, or 14%, had a grain localized to the region 11q21-22.


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

REFERENCES

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

2. Tsui LC, Buchwald M, Barker D, Braman JC, Knowlton RG, Schumm JW, Eiberg H, Mohr J, Kennedy D, Plavsic N, Zsiga M, Markiewicz D, Akots G, Brown V, Helms C, Gravius T, Parker C, Rediker K, Donis-Keller H (1985) Cystic fibrosis locus defined by genetically linked polymorphic DNA marker. Science 230:1054-1057.

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

4. White R, Woodward S, Leppert M, O'Connell P, Hoff M, Herbst J, Lalouel J-M, Dean M, Vande Woude G (1985) A closely linked genetic marker for cystic fibrosis. Nature 313:382-384.

5. Davies KE, Speer A, Herrmann F, Spiegler AWJ, McGlade S Hofker MH, Briand P, Hanke R, Schwartz M, Steinbicker V, Szibor R, Korner H, Sommer D, Pearson PL, Choutelle Ch (1985) Human X chromosome markers and Duchenne muscular dystrophy. Nucl Acids Res 13:3419-3426.

6. Monoco AP, Neve RL, Collett-Feener C, Bertelson CJ, Kurnit DM, Kunkel LM (1986) Isolation of candidate cDNAs for portions of the Duchenne Muscular Dystrophy gene. Nature 323:646-650.

7. Royer-Pokora B, Kunkel LM, Monaco AP, Goff SC, Newburger PE, Baehner RL, Sessions-Cole F, Curnutte JT, Orkin SH (1986) Cloning the gene for an inherited human disorder - chronic granulomatous disease - on the basis of its chromosomal location. Nature 322:32-38.

8. Friend SH, Bernards R, Rogel S, Weinberg R, Rapaport JM, Albert DM, Dryja TP (1986) A human DNA segment with properties of the gene that predisposes to retinoblastoma and osteosarcoma. Nature 323:643-646.

9. St George-Hyslop PH, Tanzi RE, Polinsky RJ, Haines JL, Nee L, Watkins PC, Myers RH, Feldman RG, Pollen D, Drachamn D, Growdon J, Bruni A, Foncin J-F, Salmon D, Frommelt P, Amaducci L, Sorbi S, Piacentini S, Stewart GD, Hobbs WJ, Conneally PM, Gusella JF (1987) The genetic defect causing familial Alzheimer's disease maps on chromosome 21. Science 235:885-890.

10. Tanzi RE, Gusella JF, Watkins PC, Bruns GAP, St George-Hyslop P, Van Keuren ML, Patterson D, Pagan S, Kurnit DM, Neve RL (1987) Amyloid β protein gene: cDNA, mRNA distribution, and genetic linkage near the Alzheimer locus. Science 235: 880-884.

11. Egeland JA, Gerhard DS, Pauls DL, Sussex JN, Kidd KK, Allen CR, Hostetter AM, Housman DE (1987) Bipolar affective disorders linked to DNA markers on chromosome 11. Nature 325:783-787.

12. White R, Leppert M, O'Connell P, Nakamura Y, Julier C, Woodward S, Silva A, Wolff R, Lathrop M, Lalouel J-M (1986) Construction of human genetic linkage maps: I. Progress and Perspectives. *In*: Cold Spring Harbor Symposia on Quantitative Biology. LI:29-38.

13. Bufton L, Mohandas TK, Magenis RE, Sheehy R, Bestwick RK, Litt M (1986a) A highly polymorphic locus on chromosome 16q revealed by a probe from a chromosome specific cosmid library. Hum Genet 74:425-431.

14. Maslen C, Mohandas TK, Magenis RE, Sheehy R, Litt M (1987) An unusual clustering of restriction site polymorphisms on chromosome 16q. In preparation.

15. Willard HF, Skolnick MH, Pearson PL, Mandel J-L (1985) Report of the Committee on Human Gene Mapping by Recombinant DNA Techniques. (8th International Workshop on Human Gene Mapping) Cytogenet Cell Genet 40:360-490.

16. Lander ES, Botstein D (1986) Mapping complex genetic traits in humans: New methods using a complete RFLP linkage map. *In*: Cold Spring Harbor Symposia on Quantitative Biology; LI:49-62.

17. Litt M, Bruns GAP, Sheehy R, Magenis RE (1986) 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.

18. Bufton L, Bruns GAP, Magenis RE, Tomar D, Shaw D, Brook D, Litt M (1986b) Four restriction fragment length polymorphisms revealed be probes from a single cosmid map to chromosome 19. Am J Hum Genet 38:447-460.

19. Jones C, Bill J, Larizza L, Pym B, Goodfellow P, Tunnacliffe A (1984) Relationships between genes on human chromosome 11 encoding cell-surface antigens. Somatic Cell Molec Genet 10:423-428.

20. Scott A, Phillips JA, Migeon BR (1979) DNA restriction endonuclease analysis for localization of human chromosome and globin genes on chromosome 11. PNAS 76:4563-4565.

21. Athwal RS and McBride OW (1977) Serial transfer of a human gene to rodent cells by sequential chromosome-mediated gene transfer. PNAS 74:2943-2947

22. Pyati J, Kucherlapati RS, Skoultchi AI (1980) Activation of humanβ-globin genes from nonerythroid leukemia cells by fusion with murine erythroleukemia cells. PNAS 77:3435-3439.

23. Feinberg AP, Vogelstein B (1983) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal Biochem 132:6-13.

24. Litt M, White R (1985) A highly polymorphic locus in human DNA revealed by cosmid derived probes. PNAS 82:6206-6210.

25. Silhavy TJ, Berman ML, Enquist LW (1984) Experiments with gene fusions. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

26. JPA101 was a gift from John P Adelman, The Institute for Advanced Biomedical Research and the Department of Microbiology, Oregon Health Sciences University; Portland, Oregon. JPA101 is a derivative of JM101, selected for resistance to infection with bacteriophage lambda.

27. Lathrop GM, Lalouel J-M, Julier C, Ott J (1984) Strategies for multilocus linkage analysis in humans. PNAS 81:3443-3446.

28. Botstein D, White R, Skolnick M, Davis R (1980) Construction of a genetic linkage map in man using restriction length polymorphisms. Am J Hum Genet 32:314-331.

29. Gilliam TC, Healey ST, MacDonald ME, Stewart GD, Wasmuth JJ, Tanzi RE, Roy JC, Gusella JF (1987) Isolation of polymorphic DNA fragments from human chromosome 4. Nucl Acids Res 15:1445-1458

30. Marx JL (1985) Putting the human genome on the map. Science 229:150-151.

IV. Discussion

At a time in scientific history when we are finally coming to grips with the enormity of the human genome, even to the point of contemplating sequencing it, we are still perplexed and sometimes overwhelmed by the diversity we find there. We haven't even begun to grasp the significance of the complexity we know exists. We can, however, take advantage of that diversity by using it to access information otherwise hidden in the genome. RFLPs communicate some of that variation, enabling us to expand our understanding of this intricate system. Construction of human genetic linkage maps provide the basis of this information network. In this thesis I have presented work which contributes significantly to the development of genetic maps for human chromosomal regions 11q and 16q. In the following discussion I will highlight some of the details and implications of the work presented in the preceding manuscripts, and provide some additional background on the significance of this work.

A recombinant DNA library specific for the long arms of human chromosomes 11 and 16 was constructed and successfully used to isolate cloned probes which reveal RFLP markers for these genomic regions. Physical and genetic data were used to provide evidence of colinearity between the probes and the genome. The loci revealed by these probes were characterized genetically by using family and population studies to determine inheritance

patterns, linkage parameters, and polymorphism information content. Molecular models are proposed for some of the representative RFLP loci (see Appendix I).

Manuscript 1

From the experiments summarized in manuscript 1, preliminary data regarding the physical nature of the locus D16S20 have been produced. Colinearity studies have provided information about the arrangement of Eco RI and Bam HI fragments in the probe insert (see Figures 1 and 2), furnishing the basis for a restriction map of the D16S20 region.

In addition, we have found that highly variable degrees of linkage disequilibrium exist between the well characterized RFLPs of the D16S20 compound locus. This has been noted for other loci (Litt and Jorde, 1986), with no consistent relationship found between the extent of disequilibrium and the physical distance between pairs of loci. The authors suggest that in certain small regions of the genome, factors such as mutation, genetic drift, and population admixture may outweigh those of recombination in the generation of linkage disequilibrium. The construction of a restriction map for the D16S20 locus would allow a determination of the physical distance between RFLPs providing additional insight into the genetic and physical bases of linkage disequilibrium between closely linked loci.

Another observation made regarding the physical nature of D16S20 is that mutations at CpG dinucleotides did not play a predominant role in the formation of the cluster of restriction site polymorphisms. Although the locus was originally detected by a screening system that takes advantage of the reportedly increased mutation rate at CpGs (Barker *et al*, 1984), it was soon discovered that the majority of the restriction sites which define polymorphisms at this locus do not include that dinucleotide. However, this finding is consistent with the projection of Wijsman (1984) who proposed a model for predicting the relative efficiencies of different restriction enzymes for detecting DNA variants that are the result of single base changes. She concluded that restriction enzymes containing CpG dinucleotides in their recognition sequences show no particular advantage in detecting RFLPs, based on observed versus predicted efficiencies. Observations reported in manuscript 1 will add significantly to the database regarding this issue. Further studies into the detailed physical nature of this genomic region may eventually allow speculation regarding the molecular mechanism(s) resulting in this unusual regional diversity.

Study of this region also provided physical and genetic evidence for the existence of loci in the cluster which are not yet characterized. Because the primary objective of characterizing highly polymorphic marker

loci has been obtained and would not be enhanced significantly by additional information, we did not analyze in detail all of the loci detected by probe 8-9 in this thesis. The existence of these uncharacterized loci is indicated by the hybridization patterns of the probe Φ8-9 to genomic Southern blots of DNAs digested with Taq I, Pst I, Bcl I, Asp 700 and Hae III. At least two variable bands are seen for all of the enzymes except Taq I, indicating the presence of one or more polymorphic loci. In the case of Taq I, the complexity of the fragment pattern has prevented determination of allelism, or even the number of loci detected. In an attempt to assign the bands to loci, Southern blots of Taq I digested DNAs from complete hydatidiform moles were used. Complete hydatidiform moles are hemizygous for the paternal genome (Szulman and Surti, 1978), and are therefore useful for detecting haploid fragment patterns. The results of this analysis were ambiguous, possibly due to a failure to resolve or detect all of the alleles for some of the Taq I loci. None of the enzymes show fragment size differences or the genotypes of panels of individuals studied consistent with an insertion/deletion mechanism, indicating that we are not seeing one of the other characterized loci with a different enzyme.

Genetic characterization of the compound locus detected by Φ8-9 demonstrates a complication that can arise in the calculation of polymorphism information

content (PIC) (Botstein *et al*, 1980). PIC is the probability that a child is informative for linkage. Calculation of the PIC for compound loci requires knowledge of the haplotype frequencies involved. For unrelated individuals, determination of complete haplotypes is possible only for those who are homozygous at all but one of the loci typed. When dealing with two-locus haplotypes, this requirement usually causes a modest decrease in sample size, especially if the heterozygosities at the individual loci are low. However, as the number of loci included in a haplotype increases, the frequency of such single heterozygotes decreases, causing severe attrition in sample size. Haplotyping of multiple heterozygotes requires family studies to elicit information about linkage phases. Generally, practicality dictates that allele and haplotype frequencies be determined from population studies, and little if any family information is available from which to make inferences. Even in situations where family genotypes are available, it is not always possible to infer haplotypes in multiply heterozygous individuals because phase cannot always be determined. The net result is that frequently the data from multiply heterozygous individuals are excluded from the calculation of PIC. This is appropriate at times when the numbers of multiple heterozygotes are small because the exclusion will not statistically affect the

outcome of the analysis (for an example, see Litt and Jorde, 1986). However, the problem is significantly amplified as the extent of heterozygosity increases. If linkage disequilibria values between pairs of loci in a compound locus are small, then the number of singly heterozygous individuals decreases with each additional locus considered in the compound locus. The actual extent of the decrease is a function of the percent of heterozygotes in the population for each individual locus. In addition, the number of multiple heterozygotes of determinable phase, and consequently inferable haplotype, will also generally decrease. These effects are inherent in the analyses of highly polymorphic loci and have the net effect of introducing bias into the PIC calculation. This bias is amplified as heterozygosity is increased and a larger number of multiply heterozygous individuals are excluded from the sample population.

Alternatively, population heterozygosity is simply the number of individuals heterozygous at one or more of the loci under consideration, expressed as a percent of the total population. Although it is not as powerful an indicator of the usefulness of a locus as PIC when the level of heterozygosity is low, the numerical values for population heterozygosity and PIC tend to become very close as heterozygosity increases. Consequently, for highly polymorphic loci, comparison of the calculated value of PIC and the observed value for population

heterozygosity may act as an indicator of bias in the population for which the analyses were done. It is important to remember however, that the PIC value is not dependent on heterozygosity alone, but also takes into account the probability that an offspring of a particular mating will be informative. Perhaps a more useful means of evaluation is the comparison of the percent heterozygosity for a sample population and the observed heterozygosity for the population subset that results after haplotypes have been inferred and individuals with undeterminable haplotypes have been excluded. If there is an appreciable difference between the values, it can be concluded that significant bias has been introduced into the sample, and that PIC may no longer be the most appropriate measure of usefulness of the marker locus.

Analysis of the locus D16S20 posed the question of which indicator of information content was most suitable for the study at hand. Given the population available for study, we found that exclusion of multiply heterozygous individuals at the compound locus decreased the population heterozygosity from 86%, to an effective heterozygosity of only 79%. It was felt that inference of haplotypes in order to calculate a PIC value introduced a significant bias into the analysis, and that population heterozygosity would be the most appropriate indicator of the usefulness of that compound locus for linkage studies.

We propose consideration of population heterozygosity values as acceptable measures of information content in situations where determination of haplotypes is impractical and the necessary data manipulation results in the introduction of significant bias resulting in a PIC value which is not truly reflective of the usefulness of that locus.

Manuscript 2

The results of this thesis also encourage speculation on the extent and distribution of polymorphisms throughout larger genomic regions. Library screening has so far resulted in the isolation of more polymorphic clones deriving from chromosome region 11q than from 16q. Since the initial selection of human clones was presumably random, the proportion of total 11q;16q clones should be indicative of the relative arm sizes represented in the library. Based on cytogenetic estimates of the physical sizes of these fragments, we conclude that the numbers of clones will be approximately equal for each chromosome arm. The proportion of polymorphic clones isolated from each arm *could* be reflective of the relative extent of polymorphism in these two genomic regions. Results to date, however, are not considered statistically significant as the sample size is small. In addition, the actual genetic constitution of the library is currently unknown and

further characterization is necessary before conclusions can be drawn.

The uneveness of the distribution of polymorphic loci which map to 11q was also unexpected. Seven polymorphic loci, six of which are characterized in this thesis, map between 11q22-qter. We estimate that this is approximately one-half of the long arm, based on cytogenetic appraisal of the physical size of the arm. There are several possible explanations which could account for this phenomenon. The distribution of clones found here may be reflective of the actual arrangement of site polymorphisms along 11q. There is also the possibility that the chromosome specific library from which the clones were isolated is not representative of the somatic cell hybrid from which it was constructed. The third possible explanation is that our observations are due to chance and do not reflect the actual distribution of RFLPs in the library, or the chromosomal regions studied.

Future work with this library will include mapping of both polymorphic and nonpolymorphic clones. This will eventually furnish a statistically significant representation of the library constitution, which will in turn provide information about the relative numbers of RFLPs on 11q and 16q, and the nature of the distribution of RFLPs along these regions.

In future library screenings, polymorphic clones will be localized early in the characterization process to establish priorities in regions that have the fewest mapped markers. We have recently obtained a panel of somatic cell hybrid lines containing different regions of chromosome 16 which will provide an alternative to *in situ* hybridizations for mapping 16q specific clones. If, after a significant number of clones have been mapped, the library displays signs of nonrandom distribution of all clones, it may be necessary to reconsider strategies for mapping the 11q and 16q regions. First, it would be important to first determine if possible the cause of the skewed library constitution in order to assess the continued usefulness of the somatic cell hybrid from which it was derived. Probes from other labs assigned to portions of the 11q and 16q regions appearing to be underrepresented in our library could be used to determine the library and somatic cell hybrid constitutions.

Additional Strategies

Experimental approaches specific for the detection of highly polymorphic loci are considered to be advantageous when constructing genetic linkage maps. Two strategies used in this project were designed to isolate highly polymorphic RFLP probes from chromosomal regions 11q and 16q.

A cosmid library, constructed from the same 11q;16q chromosome specific material as the phage library reported here, was initially used to isolate cosmid clones identifying highly polymorphic marker loci (Bufton *et al.*, 1986). Cosmids were initially chosen as the vector because their cloning capacity (35-50 kb) would allow easy screening of relatively large genomic fragments, increasing the potential for detecting multiple closely linked RFLP loci (Litt and White, 1985). Unfortunately, although several highly polymorphic cosmid clones were isolated, an instability of the clones, apparently inherent in many cosmid libraries, resulted in the rapid and irretrievable loss of clones. Evidence suggested that the instability of the cosmid clones was due to spontaneous deletions of part or all of the insert, resulting in cosmids too small to be packaged into phage particles for propagation. Since then, Fernandez *et al* (1986) have reported the discovery of a spontaneous insertion of the *E. coli* element, IS1, into a cosmid vector used to construct human genomic libraries. IS (insertion) elements act as hot spots for deletion events, and may explain the cosmid instability. Although it is unknown if this has occurred in the 11q;16q cosmid library, the high level of instability made it impractical as a tool for the construction of linkage maps.

Using a second strategy employed to detect highly variable loci, the 11q;16q phage library was screened for the presence of variable number tandem repeats (VNTR). VTNRs occur in several regions throughout the genome and can serve as highly polymorphic marker loci (Nakamura *et al.*, 1987). Four different synthesized oligomeric DNA probes corresponding to tandem repeat sequences involved in different insertion/deletion polymorphisms were used to screen approximately 2 x 10^4 recombinant phage clones in an attempt to detect VNTR markers on 11q or 16q. The probes were provided by R. White (Howard Hughes Medical Institute; Salt Lake City, Utah), and were synthesized based on the consensus sequences for the insulin gene associated hypervariable region (Bell *et al.*, 1982), the myoglobin minisatellite (Jeffreys *et al.*, 1985), the c-Harvey-ras-1 oncogene repeat (Capon *et al.*, 1983), and the $\psi\zeta$-globin repeat (Goodbourn *et al.*, 1983). Three positive clones of the 2 x 10^4 phage screened were obtained, but none showed polymorphisms when used to probe genomic Southern blots. At the same time, Ray White's group was using these probes to screen genomic phage libraries (P.O'Connell, pers. comm.). They found about 50 positive clones/2.5 x 10^5 recombinant phage, none of which revealed VNTR polymorphisms on Southern blot analysis (Nakamura *et al.*, 1987). From this data, they concluded that although positive clones were being detected in genomic and chromosome specific libraries, at

approximately 50 clones/genome equivalent, VNTR sequences were not generally present in phage libraries. It is now felt that because phage with inserted repetitive sequences do not grow well in *rec* A^+ bacterial hosts (Wyman *et al.*, 1985), that recombinant phage containing these tandem repeat sequences lose or delete portions or all of the insert early in propagation (Nakamura *et al.*, 1987). Since the selective mechanism inherent in the vector used to construct the 11q;16q chromosome specific phage library requires that it be propagated in a *rec* A^+ bacterial host, we decided to abandon this method of library screening in preference for more productive strategies.

Since that time, Nakamura *et al.*, have used these and seven other consensus sequence probes to successfully screen a genomic cosmid library which was grown in a *rec* A^- bacterial host (Nakamura *et al.*, 1987). They have screened 7.5×10^5 recombinant cosmid clones, isolating 77 VNTR markers with population heterozygosities between 50 and 90%. VTNRs are expected to serve as valuable markers around which complete linkage maps can be centered.

Current Applications for 11q Probes

All probes generated in this thesis are available to other researchers on a collaborative basis. The six probes which map to 11q are currently in use as part of three separate studies. In an attempt to map the gene

for neurofibromatosis, an autosomal dominant disorder which occurs in approximately one in 2500-3000 live births, researchers have eliminated many genomic regions from consideration through linkage analyses. Chromosomal region 11q is one area of the genome which has not been eliminated. The RFLP markers for 11q reported here provide a preliminary panel of probes for the assessment of linkage to the approximate distal one-half of 11q. David Barker, Robert Erickson and their coworkers are currently using these probes in an attempt to map the gene for neurofibromatosis.

The probes which map to 11q are also being used in studies investigating putative tumor suppression activity on chromosome 11. Recent reports by Stanbridge *et al* (1987) and Weissman *et al* (1987) indicate that tumor suppressor genes or anti-oncogenes exist in at least two regions of chromosome 11. Weissman *et al* (1987) have found that introduction of a normal chromosome 11 into a Wilm's tumor cell line via microcell transfer results in suppression of the tumorigenic activity of the cells. Deletions at 11p13 are associated with the development of Wilm's tumor in some cases (Yunis and Ramsey, 1980). Microcell transfer of a translocated human chromosome, 11pter>11q23::Xq26>Xqter, with the terminal portion of the long arm of 11 replaced by Xq26-Xqter, suppressed tumorigenicity in the Wilm's tumor cell line suggesting

the presence of a anti-oncogene, possibly in the associated region of deletion.

In similar studies, Stanbridge *et al* (1987) are using microcell transfer to introduce portions of chromosome region 11q into HeLa cells. Introduction of the translocated chromosome used by Weissman *et al* in the Wilm's tumor study suppresses the tumorigenic expression of HeLa cells. Preliminary studies indicate that a gene on mid-11q may be responsible for this activity (E. Stanbridge, pers. comm.). The RFLP probes identified here are being used to delineate regions on chromosome 11 involved in the control of tumorigenic expression of some cell types.

CEPH Typing and Construction of a Map of 11q

The six 11q specific polymorphic loci described in this thesis are also being used by the Ray White lab to type the CEPH families as the basis for constructing a complete linkage map of 11q. Table 1 shows these and a recent listing of additional published loci deriving from 11q, which will contribute to the genetic map. Multipoint linkage analysis using these loci and the CEPH reference families will indicate order and estimate genetic distances between these markers, culminating in a genetic linkage map for 11q.

Figure 1. Eco RI (E) digestion of Φ8-9 produces insert-derived fragments of 11.5, 4.0, and 1.8 kb. Southern blot analysis of Eco RI digested genomic DNA samples probed with Φ8-9 reveals genomic derived fragments of >20, 12.5, and 1.8 kb. In both cases there is a 1.8 kb fragment, indicating that it must occur as the center fragment, flanked by the two larger fragments as seen below. The bands detected on Southern blot analysis are of equal or larger size than the probe derived fragments, because the vector associated ends of the Eco RI digested insert fragments were cleaved with Sau 3A during the initial cloning procedure.

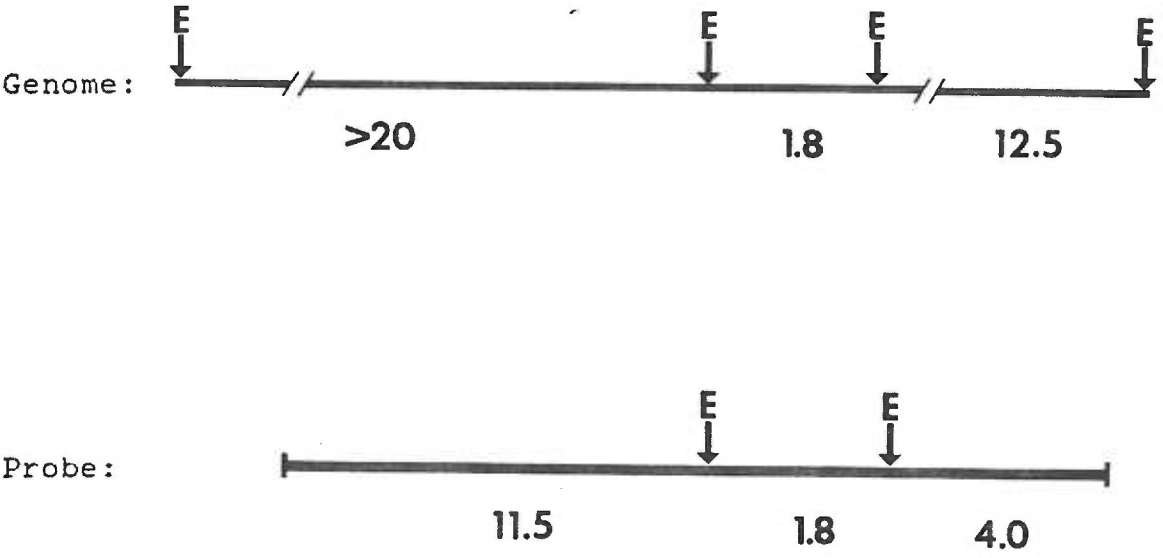


Figure 2. The order of Bam HI sites (B) is determined in an analysis seen with Bam HI. Digestion of 8-9 with Bam HI produces insert derived fragments of 8.3, 5.5, 3.5 kb. Southern blot analysis of genomic DNA samples probed with 8-9 reveals three genomic bands of 13.0, 5.5 and 4.9 kb. This indicates that the 13.0 and 4.9 kb fragments must flank the 5.5 kb piece, as seen below.

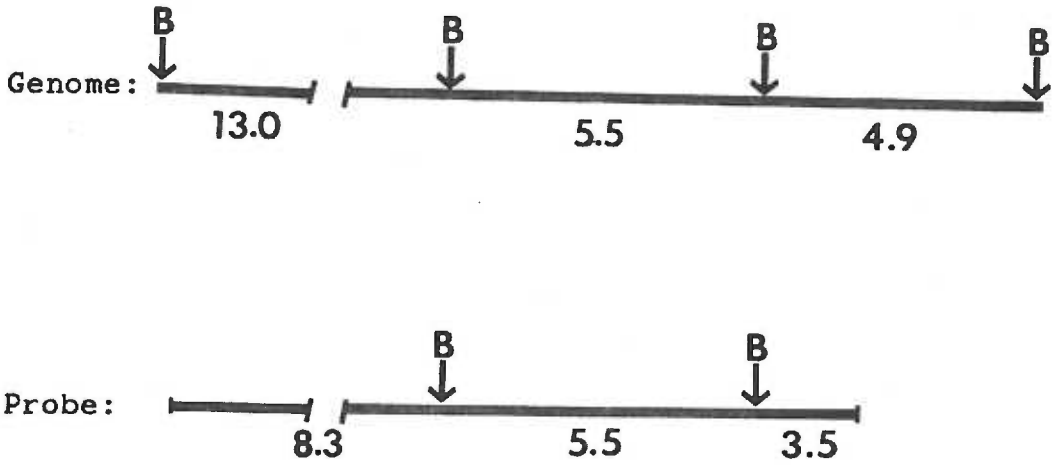


Table 1

Elements for a Primary Linkage Map of 11q

Locus	Map Location	PIC
[1]D11S15	11	0.36
[1]D11S16	11	>0.54
[2]PYGM	11q	?
[1]APOA1	11q13	0.50
[1]APOC3	11q13	?
[1]PGA	11q13	?
[3]INT2	11q13	>0.37
[1]APOA4	11q13	?
[4]D11S.3.7	11q13-q14	0.37
[1]D11S23	11q13-q23	0.18
[1]*D11S24	11q13-q23	0.24
*D11S35	11q22	0.58
*D11S36	11q22	0.46
*D11S84	11q22	0.32
[5]D11S85	11q22-23	0.37
[5]D11S29	11q23	?
[1]ETS1	11q23-q24	0.26
[6]*PPD	11q23-qter	0.37
*D11S34	11q23-qter	0.35
D11S83	11q23-qter	0.38

*Manuscript 2, this thesis

[1]Willard *et al*, 1985
[2]R.Fletterick (pers. Comm.)
[3]Casey *et al.*, 1986
[4]Dietzsch *et al.*, 1987
[5]Warnich *et al.*, 1986
[6]Llewellyn *et al.*, 1987

V. APPENDIX I

Figure 1. Probe Φ2-25 (-----) detects three variable fragments with Msp I (M) of 4.35 kb and cosegregant 3.2 + 1.15 kb. Consequently, the probe must hybridize to the region of the variable restriction site (*). Other probes detecting three variable fragments at a single locus where the sizes of the two smaller fragments add to the size of the largest fragment fit a similar model.

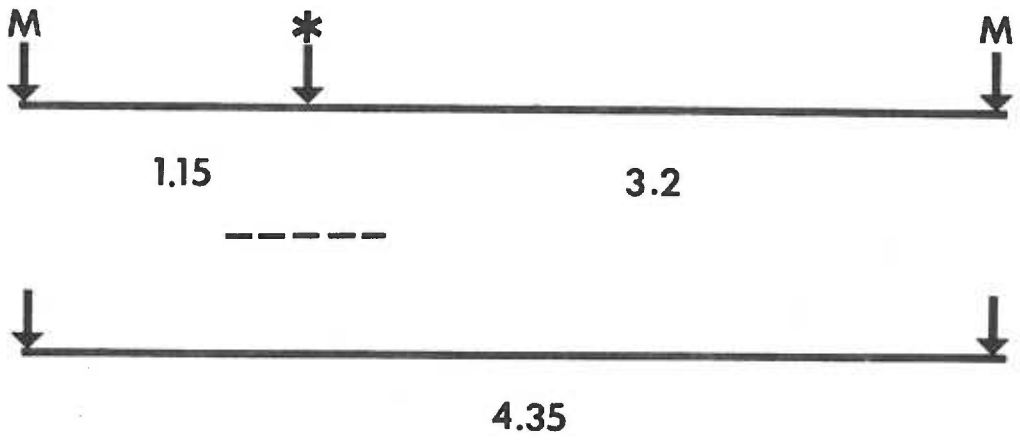


Figure 2. Probe p2-7-1D6 (-----) detects two variable Taq I fragments of 6.4 and 4.25 kb. A separate subclone, p2-7-2A7 detects the 6.4 and the 2.05 kb fragments. (T) = Taq I site, * = variable site.

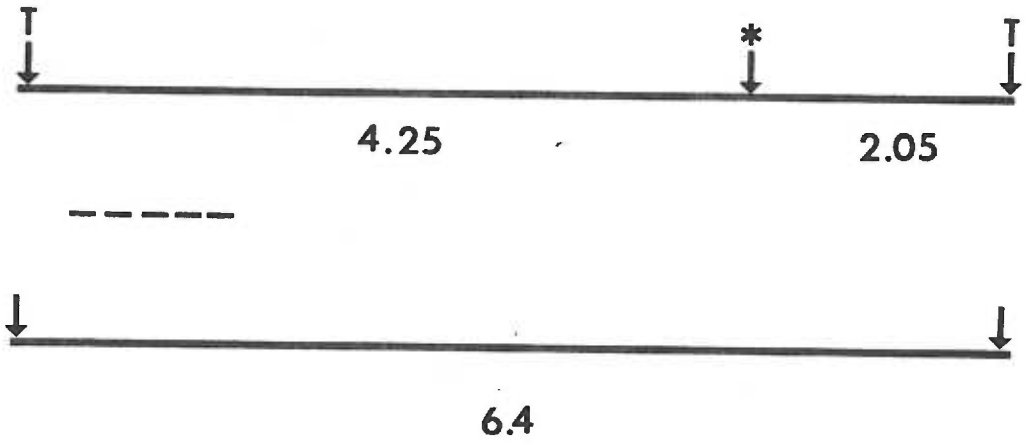


Figure 3. Probe 02-14 (-----) detects two loci, one with Rsa I (R) and one with Pvu II (P). Rsa I reveals 2.95 and 2.1 kb variable bands. Pvu II detects 12.5 and 11.5 kb variable bands. * = variable restriction site.

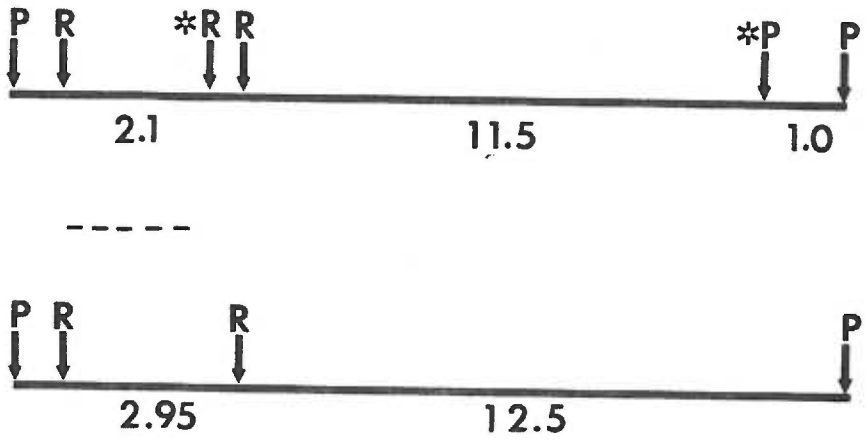
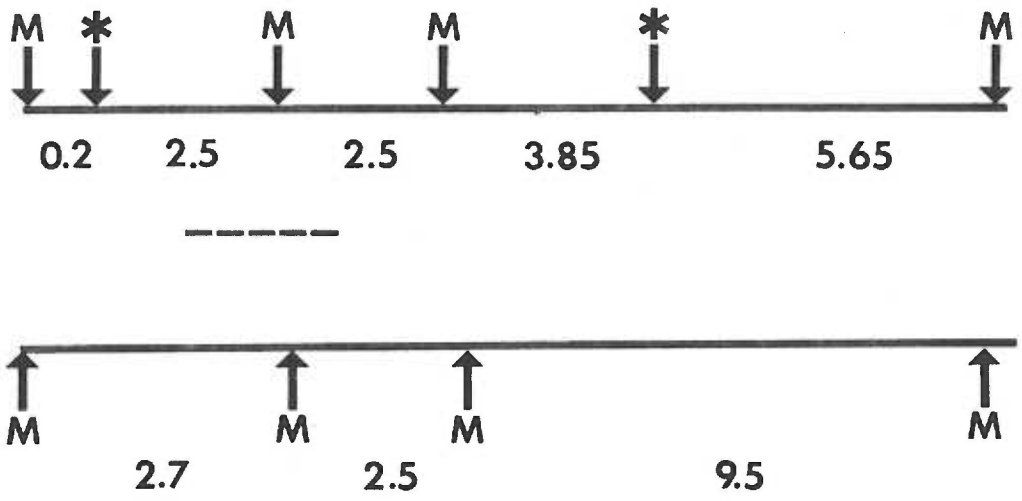


Figure 4. Probe Φ6-3 (-----) detects two loci with a single enzyme, Msp I (M). Four variable fragments, 9.5, 3.85, 2.7, 2.5 kb, and a single 2.5 kb constant band are revealed. * = variable restriction site.



VI. APPENDIX II

Technical Acknowledgements

Gail Bruns (Children's Hospital, Boston MA) provided somatic cell hybrid mapping panels used in manuscripts 1 and 2.

Christopher Dubay (Medical Genetics, OHSU) provided computer analysis using the LINKAGE program of data shown in Table 4 of manuscript 2.

Tom Glaser (MIT) provided somatic cell hybrid lines MC-1, TGD5D1-1, and G1-7 used for 11q specific mapping in manuscript 2.

Carol Jones (U. of Colorado) provided somatic cell hybrid lines J1-11 and J1-44 used for 11q specific mapping in manuscript 2.

Ruth Litt (Biochemistry, OHSU) produced many of the Southern blots used in manuscripts 1 and 2.

TK Mohandas (UCLA) provided the somatic cell hybrid containing the 11q;16q human translocation used to create the phage library on which manuscripts 1 and 2 were based.

Robert Sheehy (CDRC, OHSU) provided the *in situ* hybridization data for manuscripts 1 and 2.

VII. REFERENCES

REFERENCES

Anand, R (1986) Pulse field gel electrophoresis: a technique for fractionating large DNA molecules. TIGS 2:278-283

Antonarakis SE, Boehm C, Giardina PJV (1982) Nonrandom association of polymorphic restriction sites in theβ-globin cluster. PNAS 79:137-141

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

Bell GI, Selby MJ, Rutter WJ (1982) The highly polymorphic region near the human insulin gene is composed of simple tandem repeated sequences. Nature 295:31-35

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

Bufton L, Mohandas TK, Magenis RE, Sheehy R, Bestwick RK, Litt M (1986) A highly polymorphic locus on chromosome 16q revealed by a probe from a chromosome specific cosmid library. Hum Genet 74:425-431

Capon DJ, Chen EY, Levinson AD, Seeburg PH, Goeddel DV (1983) Complete nucleotide sequences of the T24 human bladder carcinoma oncogene and its normal homologue. Nature 302:33-37

Casey G, Smith R, McGillivaray D, Peters G, Dickson C (1986) Characterization and chromosome assignment of the human homolog of Int-2, a potential proto-oncogene. Mol Cell Biol 6:502-510

Cavalli-Sforza LL, Kidd JR, Kidd KK, Bucci C, Bowcock AM, Hewlett BS, Friedlaenders JS (1987) DNA markers and genetic variation in the human species. Cold Spring Harbor Symposia on Quantitative Biology LI:411-417, Cold Spring Harbor, NY

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

Chakravarti A, Buetow KH, Antonarakis Se, Waber PG, Boehm CD, Kazazian HH (1984) Nonuniform recombination within the human α-globin gene cluster. Am J Hum Genet 36:1239-1258

Collins FS, Drumm ML, Cole JL, Lockwood WK, Vande Woude GF, Iannuzzi MC (1987) Construction of a general human chromosome jumping library with applications to cystic fibrosis. Science 235:1046-1049

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

Detera-Wadleigh D, Berrettini WH, Goldin LR, Boorman D, Anderson S, Gershon ES (1987) Close linkage of c-Harvey-ras-1 and the insulin gene to affective disorder is ruled out in three North American pedigrees. Nature 325:806-808

Dietzsch E, Retief AE, Fricke J, du Plessis L, Nicholson DL, Kotze MJ (1987) Frequent RFLP recognized by an anonymous sequence localized to 11q13-14

Donahue RP, Bias WB, Renwick JH, McKusick VA (1968) Probable assignment of the Duffy blood group locus to chromosome 1 in man. PNAS 61:949-955

Dracopoli NC, Houghton AN, Old LJ (1985) Loss of polymorphic restriction fragments in malignant melanoma: implications for tumor heterogeneity. PNAS 82:1470-1474

Egeland JA, Gerhard DS, Pauls DL, Sussex JN, Kidd KK, Allen CR, Hostetter AM, Housman DE (1987) Bipolar affective disorders linked to DNA markers of chromosome 11. Nature 325:783-787

Fearon ER, Feinberg AP, Hamilton SH, Vogelstein B (1985) Loss of genes on the short arm of chromosome 11 in bladder cancer. Nature 318:377-380

Feinberg AP and Vogelstein B, (1983) A technique for radiolabeling DNA restriction fragments to high specific activity. Anal Biochem 132:6-13

Fernandez C, Larhammar D, Servenius B, Rask L, Peterson PA (1986) Spontaneous insertions into a cosmid vector: a warning. Gene 42:215-219

Francomano CA and Kazazian HH (1986) DNA analysis in genetic disorders. Ann Rev Med 37:377-395

Friend SH, Bernards R, Rogelj S, Weinberg RA, Rapaport JM, Albert DM, Dryja TP (1986) A human DNA segment with properties of the gene that predisposes to retinoblastoma and osteosarcoma. Nature 323:643-646

Frischauf A-M, Lehrach H, Poutska A, Murray N (1983) Lambda replacement vectors carrying polylinker sequences. J Mol Biol. 170:827-842

Goldgaber D, Lerman MI, McBride OW, Saffiotti U, Gajdusek DC (1987) Characterization and chromosomal localization of a cDNA encoding brain amyloid of Alzheimer's disease. Science 235:877-880

Goodbourne SEY, Higgs DR, Clegg JB, Weatherall DS (1983) Molecular basis of length polymorphisms in the human zeta-globin gene complex. PNAS 80:5022-5026

Goodfellow P, Davies K, Ropers HH (1985) Report of the committee on the genetic constitution of the X and Y chromosomes. Cytogenet Cell Genet 40:296-352

Gusella JF (1986) DNA polymorphism and human disease. Ann Rev Biochem 55:831-854

Gusella JF, Keys C, Varsanyi-Breiner A, Kao F-T, Jones C, Puck TT, Housman D (1980) Isolation and localization of DNA segments from specific human chromosomes. PNAS 77:2829-2833

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 (1983) A polymorphic DNA marker genetically linked to Huntington's disease. Nature 306:234-238

Hadfield, C (1983) Chromosome Walking. BRL Focus 5(4):1-5

Haldane JBS, Smith CAB (1947) A new estimate of linkage between the genes for colour-blindness and haemophilia in man. Ann Eugen 14:10-15

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

Jabs EW, Wolf SF, Migeon BR (1984) Characterization of a cloned DNA sequence that is present at centromeres of all human autosomes and the X chromosome and shows polymorphic variation. PNAS 81:4884-4888

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

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

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

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

Kan YW and Dozy A (1978) Antenatal diagnosis of sickle-cell anemia by DNA analysis of amniotic-fluid cells. Lancet 2:910-912

Klein G and Klein E (1985) Evolution of tumours and the impact of molecular oncology. Nature 315:190-195

Knowlton R, Schumm J, Braman J, Barker D, Donis-Keller H (1985) Relative efficiencies of restriction endonucleases in detecting restriction fragment length polymorphisms. Cytogenet Cell Genet 40:670

Koufos A, Hansen MF, Copeland NG, Jenkins NA, Lampkin BC, Cavenee WK (1985) Loss of heterozygosity in three embryonal tumours suggests a common pathogenetic mechanism. Nature 316:330-334

Krontiris TG, DiMartino NA, Colb M, Parkinson DR (1985) Unique allelic restriction fragments of the human Ha-ras locus in leukocyte and tumor DNAs of cancer patients. Nature 313:369-373

Lander ES, Botstein D (1987) Mapping complex genetic traits in humans: New methods using a complete RFLP linkage map. Cold Spring Harbor Symposia on Quantitative Biology LI:49-62, Cold Spring Harbor, NY

Larson RA, Williams SF, Le Beau MM, Bitter MA, Vardiman JW, Rowley JD (1986) Acute myelomonocytic leukemia with abnormal eosinophils and inv(16) or t(16;16) has a favorable prognosis. Blood 68:1242-1249

Lathrop GM, Lalouel J-M, Julier C, Ott J (1984) Strategies for multilocus linkage analysis in humans. PNAS 81:3443-3446

Lathrop GM, Lalouel J-M, Julier C, Ott J (1985) Multilocus linkage analysis in humans: Detection of linkage and estimation of recombination. Am J Hum Genet 37:482-498

Lathrop GM, Lalouel J-M, White RL (1986) Construction of human genetic linkage maps: Likelihood calculations for multilocus linkage analysis. Genet Epidemiol 3:39-52

Lewin B (1983) Genes. John Wiley and Sons Inc.; NY, NY

Lewontin RC (1967) Population genetics. Ann Rev Genet 1:37-70

Lewontin RC (1973) Population genetics. Ann Rev Genet 7:1-17

Lewontin RC (1986) Population genetics. Ann Rev Genet 19:81-102

Litt M, Bruns GAP, Sheehy R, Magenis RE (1986) 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

Litt M and Jorde LB (1986) Linkage disequilibrium between pairs of loci within a highly polymorphic region of chromosome 2q. Am J Hum Genet 39:166-178

Litt M and White RL (1985) A highly polymorphic locus in human DNA revealed by cosmid derived probes. PNAS 82:6206-6210

Little P (1986) Finding the defective gene. Nature 321:558-559

Llewellyn DH, Kalsheker NA, Elder GH, Harrion PR, Chretien S, Goossen SM (1987) A Msp I polymorphism for the human porphobilinogen deaminase gene. NAR 15:1349

Maniatis T, Fritsch EF, Sambrook J (1982) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory; Cold Spring Harbor, New York

Maniatis T, Hardison R, Lacy E, Lauer J, O'connell C, Quon D (1978) The isolation of structual genes from libraries of eukaryotic DNA. Cell 15:687-701

Marx JL (1985) Putting the human genome on the map. Science 229:150-151

Monaco AP, Neve RL, Colletti-Freener C, Bertelson CJ, Kurnit DM, Kunkel LM (1986) Isolation of candidate cDNAs for portions of the Duchenne muscular dystrophy gene. Nature 323:646-650

Morton N (1955) Sequential tests for the detection of linkage. Am J Hum Genet 7:277-281

Morton NE (1962) Segregation and linkage. Methodology in Human Genetics Holden Day, San Francisco

Nakamura Y, Leppert M, O'Connell P, Holm T, Culver M, Martin C, White RL (1987) Highly polymorphic DNA markers for human gene mapping. Am J Hum Genet In Press

Nakamura Y, Leppert M, O'Connell P, Wolff R, Holm T, Culver M, Martin C, Fujimoto E, Hoff M, Kumlin E, White R (1987) Variable number tandem repeat (VNTR) markers for human gene mapping. Science 235:1616-1622

Orkin SH (1987) Reverse genetics and human disease. Cell 47:845-850

Proudfoot NJ, Gil A, Maniatis T (1982) The structure of the human zeta-globin gene and a closely linked, nearly identical pseudogene. Cell 31:553-563

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

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

Rovigatti U, Watson DK, Yunis JJ (1986) Amplification and rearrangement of Hu-ets-1 in leukemia and lymphoma with involvement of 11q23. Science 232:398-400

Royer-Pokora B, Kunkel LM, Monaco AP, Goff SC, Newburger PE, Baehner RL, Sessions-Cole F, Curnutte JT, Orkin SH (1986) Cloning the gene for chronic granulomatous disease on the basis of chromosomal location. Nature 322:32-38

Saxon PJ, Srivatsan ES, Stanbridge EJ (1987) Introduction of human chromosome 11 via microcell transfer controls tumorigenic expression of HeLa cells. EMBO J 5:3461-3466

Southern EM (1975) Detection of specific DNA fragments separated by gel electrophoresis. J Mol Biol 98:503-517

St George-Hyslop P, Tanzi RE, Polinsky RJ, Haines JL, Nee L, Watkins PC, Myers RH, Feldman RG, Pollen D, Drachman D, Growdon J, Bruni A, Foncin J-F, Salmon D, Frommelt P, Amaducci L, Sorbi S, Piacentini S, Stewart GD, Hobbs WJ, Conneally PM, Gusella JF (1987) The genetic defect causing familial Alzheimer's disease maps on chromosome 21. Science 235:885-890

Stewart GD, Harris P, Galt J, Ferguson-Smith MA (1985) Cloned DNA probes regionally mapped to human chromosome 21 and their use in determining the origin of nondisjunction. Nucleic Acids Res 13:4125-4132

Tanzi RE, Gusella JF, Watkins PC, Bruns GAP, St George-Hyslop P, Van Keuren ML, Patterson D, Pagan S, Kurnit DM, Neve RL (1987) Amyloid β protein gene: cDNA, mRNA distribution, and genetic linkage near the Alzheimer locus. Science 235:880-884

Thompson E (1984) Information gain in joint linkage analysis. IMAJ Math Appl Med Biol 1:31-49

van der Ploeg LHT, Flavell RA (1980) DNA methylation in the humanβ-globin locus in erythroid and nonerythroid tissues. Cell 19:947-958

van Ommen GJB, Verkerk JMH, Hofker MH, Monaco AP, Kunkel LM, Ray P, Worton R, Wieringa B, Bakker E, Pearson PL (1986) A physical map of 4 million bp around the Duchenne muscular dystrophy gene on the human X-chromosome. Cell 47:499-504

Vogelstein B, Fearon EF, Hamilton SR, Feinberg AP (1985) Use of restriction fragment length polymorphisms to determine the clonal origin of human tumors. Science 227:642-645

Warnich L, Kotze MJ, Retief AE, Dietzsch E, Fox MC, Kotze GM, Nicholson DL, Retief E, Oosthuizen CJJ (1986) An anonymous human single copy genomic clone, D11S29 at 11q23, identified by a moderately frequent RFLP. NAR 14:1920

Weissman BE, Saxon PJ, Pasquale SR, Jones GR, Geiser AG, Stanbridge EJ (1987) Introduction of a normal human chromosome 11 into a Wilm's tumor cell line controls its tumorigenic expression. Science 236:175-180

White R, Leppert M, O'Connell P, Nakamura Y, Julier C, Woodward S, Silva A, Wolff R, Lathrop M, Lalouel J-M (1987) Construction of human genetic linkage maps: Progress and perspectives. Cold Spring Harbor Symposia on Quantitative Biology, LI:29-38, Cold Spring Harbor, NY

Wijsman EM (1984) Optimizing selection of restriction enzymes in the search for DNA variants. Nucl Acids Res 12:9209-9226

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

Wyman AR and White RL (1980) A highly polymorphic locus in human DNA. PNAS 77:6754-6756

Wyman AR, Wolffe LB, Botstein D (1985) Propagation of some human DNA sequences in bacteriophage lambda vectors requires mutant *Escherichia coli* hosts. PNAS 82:2880-2884

Youssoufian H, Kazazian HH, Phillips DG, Aronis S, Tsiftis G, Brown VA, Antonarakis SE (1986) Recurrent mutations in haemophilia A give evidence for CpG mutation hotspots. Nature 324:380-382

Yunis JJ, Ramsey NKC (1980) Familial occurrence of the aniridia-Wilm's tumor syndrome with deletion 11p13-14.1. J Peds 96:1027-1030