# A HIGH-RESOLUTION GENETIC MAP OF HUMAN CHROMOSOME 16 AND LOCALIZATION OF THE *MEF* GENE

A thesis submitted to the University of Adelaide
for the Degree of Doctor of Philosophy

by

YANG SHEN M.S.

Department of Paediatrics, Women's and Children's Hospital,
The University of Adelaide

February, 1994

# DECLARATION

The work described in this thesis was performed by the candidate except where acknowledged. This thesis contains no material has been accepted for the award of any other degree or diploma in any University. To the best of my knowledge and belief, the thesis contains no material previously or written by another person, except where due reference is made in the text of the thesis.

I consent to this thesis being made available for photocopying and loan if accept for the award of the degree of Ph. D .

Yang Shen

# TABLE OF CONTENTS

# ACKNOWLEDGMENTS

This project was carried out in the Department of Cytogenetics and Molecular Genetics at the Women's and Children's Hospital from December of 1990 to February of 1994. It has been a great pleasure to work on the construction of a PCR-based genetic map of human chromosome 16 and localization of the *MEF* gene in the Department.

I wish to thank Professor Grant R. Sutherland, my supervisor, for providing me the opportunity to study in this Department, for his supervision, encouragement, and advice during the course of this study and for his helpful comments in reviewing the drafts of this thesis.

I sincerely thank Dr. Robert I. Richards, my supervisor, for his valuable direction, encouragement, advice, discussion and suggestion throughout the course of this study and for his helpful advice and comments during preparation of this thesis.

I am grateful to Helen Kozman for doing the linkage analysis and for comments in the drafts of this thesis; to Drs David F. Callen and John C. Mulley, for their advice, discussion and comments during the course of this study and in reviewing the drafts of this thesis.

I thank Andrew Thompson, Hilary Phillips, Liang-Zhong Chen, Sinoula Apostolou and Scott Whitmore for their help with some of genotyping, physical mapping, to Julie Nancarrow for helping with automated sequencing, to Kathy Holman for oligo synthesizing, to Sharon Lane for making hybrid cell lines and providing many hours of proof reading the manuscript, to Jean Spence for extracting hybrid DNAs, to Agi Gedeon for encouragement and comments in

preparation of this thesis, to Andrew Donnelly for encouragement, to Sui Yu for encouragement and the thesis tips and to Roland Hermanis for the photography.

Many thanks to all the colleagues for creating an enthusiastic and friendly working atmosphere. I particularly thank Kathie Friend and Kathy Holman for their assistance in various ways and valuable friendship.

I am deeply indebted to my husband, Zheng-Tai Liu, for his constant support, understanding and encouragement through my study, and to my son, Kai-Xi Liu, with whom I have had many enjoyable times while studying for my Ph. D. I am also very grateful to my parents, who are in P. R. China, for their constant support, encouragement through my study.

The financial assistance provided by the Department of Cytogenetics and Molecular Genetics in the Women's and Children's Hospital is gratefully acknowledged.

SUMMARY

Genetic linkage maps of human chromosomes are important tools for the localization of disease genes. One of the first five year goals of the Human Genome project is the construction of genetic maps of all human chromosomes with highly polymorphic markers spaced an average of 2-5 centimorgans apart (Jordan, 1992). A number of genetic maps have been constructed for human chromosome 16 (Donis-Keller et al., 1987; Keith et al., 1990; Julier et al., 1990; NIH/CEPH collaborative mapping group, 1992; Kozman et al., 1993). These maps were based mainly on RFLP markers which are less efficient for localization of disease genes and for refining linkage distance for positional cloning than highly polymorphic PCR based markers. Therefore, there was a need for more highly informative markers to be placed on high density cytogenetic-based physical and genetic linkage maps of chromosome 16.

Two approaches were used to isolate simple tandem repeat (STR) markers. Initially the random isolation approach was exploited to generate STR markers from the chromosome 16 cosmid library (Stallings et al., 1990). After (AC)n repeats isolated from the random approach were mapped to 9 intervals of chromosome 16 defined by breakpoints in somatic cell hybrids, a second targetted approach was used to generate additional repeat markers with which to fill in the deficient intervals or to isolate more repeat markers in the regions of particular interest [fragile sites *FRA16A* and *FRA16B*, disease gene regions of adult autosomal polycystic kidney disease (PKD1), familial

Mediterranean fever (FMF) and Batten disease (CLN3)]. This second approach was involved the isolation of STR markers from cosmid clones which have been previously mapped to specific regions of chromosome 16 (Callen et al., 1992; Stallings et al., 1992).

A total of 32 (AC)n repeat markers were isolated from these two approaches. 22 (AC)n repeat markers had heterozygosity greater than 0.68. The 32 (AC)n repeat markers and 32 (AC)n repeat markers from other laboratories were physically mapped on the high-resolution cytogenetic map of chromosome 16 using the hybrid panel. 79 STR markers were used for construction of a genetic linkage map of human chromosome 16 (Shen et al., submitted). Of these 27 markers were generated by the candidate, 20 other STR markers were isolated by other members of our laboratory and 32 STR markers were from other laboratories

This PCR-based genetic linkage map of human chromosome 16 was constructed from 79 STR markers, 1 VNTR marker and 1 RFLP marker. These 2 non-STR markers were chosen because they extend the map towards the telomeres. This map covers the entire length of chromosome 16. The length of the map is in remarkable agreement with those published genetic and chiasma maps of chromosome 16 (Shen et al, submitted). The median distance and the average distance between markers on the framework map is 2.7 and 3.2 cM, respectively. In comparison with the averaged resolutions of other STR-based linkage maps of human chromosome 1, 4, 11, 12, 13, 18, 20 and 22, the resolution of this map is much higher.

The framework and comprehensive maps were anchored to the high-resolution cytogenetic map, which was divided into 66 breakpoint intervals (on average 1.5 Mb per interval) by a panel of 67 hybrids. The cytogenetic map is one of the most detailed maps available for any of the autosomes (Callen et al., 1992). It is apparent that the combination of genetic linkage analysis and physical mapping can be extremely helpful in resolving locus order at the resolution of the comprehensive map.

These integrated genetic and physical maps of human chromosome 16 (Shen et al. submitted) provided an efficient means for regional localization of genetic disorders located on chromosome 16, for detection of loss of heterozygosity in cancers and imprinting of chromosomes in inherited disorders, for evaluation of linkage disequilibium and disease causing mutations, and for analysis of multifactorial diseases.

To localize the gene (*MEF*) responsible for familial Mediterranean fever, which is an autosomal recessive disorder characterized by attacks of fever and serosal inflammation, 4 (AC)n repeat markers (*D16S291, D16S94, D16S523* and *D16S453*) were genotyped on 62 FMF families. The linkage analysis was carried out using the computer program MLINK of LINKAGE (version 5.1). The lod scores showed these 4 markers were significantly linked to *MEF*. The observed recombination events, homozygosity mapping and multipoint linkage analysis defined the centromeric boundary at *D16S523* and the telomeric boundary at *D16S246*. The *MEF* gene was localized between these two markers in a genetic interval of less than 1.6 cM.

# CHAPTER 1

## LITERATURE REVIEW

## 1.1. INTRODUCTION

The genetic variation present in humans represents an invaluable resource for application to molecular and medical genetics. The direct cloning and characterization of genes underlying human inherited traits has helped elucidate the molecular basis of a wide range of physiological processes in higher organisms. The difficulty with extending this approach to all heritable traits and diseases is that the vast majority of such genes are known only by their phenotype, however their biochemical basis is unknown. The positional cloning of such genes requires the identification of their location in the genome determined by a combination of genetic linkage mapping and the development of a contig of large genomic fragments across the region. The initial chromosomal assignment is dependent on the availability of useful genetic markers positioned beforehand on a reference map.

Since the potential of restriction fragment length polymorphisms (RFLPs) was realised in 1980 (Botstein et al., 1980), they have been utilised for construction of the first genetic linkage map of the human genome (Donis-Keller et al., 1987) and for the initial localization of disease genes, such as Duchenne muscular dystrophy (Davies et al., 1983), Huntington disease (Gusella et al., 1983), cystic fibrosis (Tsui et al., 1985), adult polycystic kidney disease (Reeders et al., 1985) and so on. Although RFLP markers increased the power of linkage studies by greatly expanding the number of available markers, it was found that their polymorphic

information content was often low, and that typing of large numbers of RFLP loci by Southern analysis was extremely labour-intensive. Thus, the application of RFLPs had limitations. Other classes of polymorphic genetic markers, variable number of tandem repeats (VNTRs, Nakamura et al., 1987; minisatellites, Jeffreys et al., 1985) located predominantly in the terminal parts of chromosomes and α satellite repetitive sequences located at centromeres (Willard et al., 1986) of human chromosomes, leave large regions of the genome unlinked to multiallelic marker loci. Although such loci were often very informative, genotyping was still based on Southern analysis. With the dawn of the 1990s, a genetic element, which contains dinucleotide repeats (AC)n, was observed with high polymorphic content, high density throughout all eukaryotic genomes and which could be easily genotyped using PCR (Weber and May, 1989; Litt and Luty, 1989; Tautz, 1989). The mono-, tri- and tetra-nucleotide repeats were also found to be polymorphic and observed in eukaryotic genomes (Tautz, 1989; Zuliani and Hobbs, 1990; Edwards et al., 1991; Beckmann and Weber, 1992). They are defined as relatively short (< 100 bp) runs of tandemly repeated DNA with repeat units of 6 bp or less (Beckmann and Weber, 1992) and they are referred to as simple tandem repeats (STRs), simple sequence repeats (SSRs) or microsatellites. Since their discovery, the highly polymorphic simple tandem repeats have become very useful markers for construction of genetic maps, localization of disease genes, and other applications involving DNA polymorphisms.

This literature review will cover the following: 1) frequency, distribution and polymorphism of STRs in eukaryotic genomes, and possible biological functions of STRs; 2) the applications of STRs in the construction of genetic background maps, genetic localization of disease genes using the background maps, detection of DNA variations and population studies; 3) published genetic maps of human chromosome 16; 4) the approaches used for the construction of genetic maps based on STR markers; and 5) the aims of the project.

## 1.2. THE PROPERTIES OF THE SIMPLE TANDEM REPEATS (STRS)

### 1.2.1. The Distributions of STRs in Eukaryotic Genomes

The simple tandem repeats containing all nucleotide combinations have been identified. The most common STRs found in mammalian genomes are (dA-dC)n·(dT-dG)n dinucleotide repeats and are often referred to as (AC)n repeats. Their presence in high copy number and their dispersion throughout the genomes of eukaryotic species was demonstrated a decade ago by Meisfeld et al.(1981) and Hamada et al.(1982b). When Meisfeld and his colleagues screened a human gene library with a mouse ribosomal gene non-transcribed spacer probe (rDNA NTS), a repetitive sequence (dT-dG)17 was found in a 251 bp segment between the human δ and β globin genes. Subsequently, the genomes of *Xenopus*, pigeon, slime mold and yeast were examined, and reiterated sequences homologous to both the mouse rDNA NTS and human globin repeat were found in every case. This repeated

sequence appears to be conserved in eukaryotic genomes. When studying Z conformation (a left-handed helical conformation) of poly(dA-dC)·poly(dT-dG) DNA sequences, Hamada and his colleagues used DNAs from *Drosophila melanogaster*, *Xenopus laevis*, *Saccharomyces cerevisiae*, salmon sperm, chicken, mouse, calf and human. After hybridization of these DNAs to poly(dA-dC)·poly(dT-dG) probe, a "smear" pattern in the *Eco*RI-digested total DNA from all species was observed. It was therefore suggested that (dA-dC)n sequences are randomly dispersed in these genomes. This result indicated the presence of many copies of the (dA-dC)n sequences in each genome. The approximate number of (dA-dC)n elements in various genomes was determined subsequently by dot blot hybridization. The approximate number of the (dA-dC)n elements in the examined genomes ranged from $1 \times 10^2$ to $2 \times 10^5$ (e.g. human, mouse, chicken, *Drosophila* and yeast have $5 \times 10^4$, $10^5$, $4 \times 10^3$, $2 \times 10^3$ and $10^2$, respectively). The organisms with a larger genome seem to have greater numbers of the (dA-dC)n elements. Their study also showed the (dA-dC)n sequences adopt the Z-DNA conformation under some conditions.

In 1986 Gross and Garrard reported (dA-dC)n sequences were not detectable in non-nuclear genomes. They investigated *Saccharomyces cerevisiae* mitochondrial DNA and the genomes of 13 phylogenetically diverse species of prokaryotes representative of both eubacteria and archaebacteria. By using a hybridization assay and searching the GenBank database, (dA-dC)n sequences were found to be absent

from all available prokaryotic genomes and the complete human, bovine and mouse mitochondrial genomes. They also examined the chromatin structure of (dA-dC)n elements in cultured mammalian cells. The results revealed evidence that these sequences do not exist to a significant extent in the Z conformation *in vivo*, instead, they appear to quantitatively adopt a distinctive, "alternating-B" conformation on the nucleosomal surface.

Pardue et al. (1987) investigated the distribution of (dA-dC) sequences in a number of *Drosophila* species by *in situ* hybridization to polytene chromosomes. The results revealed a clearly non-random distribution of chromosomal sites for this sequence. Sites are distributed over most euchromatin regions but are absent from the β-heterochromatin of the chromocenter. The specific site distributions are conserved in different species, including distantly related *Drosophila* species. It implied that the evolutionary conservation of the distribution of (dA-dC)n sequences is of functional importance.

Stallings et al. (1991) also investigated the distribution and conservation of (dA-dC)n sequences in several mammalian genomes. The distribution of (dA-dC)n sequences in the human genome was determined by the analysis of over 8,000 cosmid clones containing human inserts DNA which came from a chromosome 16-specific cosmid library (Stallings et al., 1990). On average, one (dA-dC)n sequence occurs every 30 kb in DNA from euchromatic regions. (dA-dC)n sequences are

significantly underrepresented in centric heterochromatin. The density of (dA-dC)n sequences in the human genome was also estimated by analyzing GenBank sequences that include introns and flanking sequences. The frequency of (dA-dC)n sequences identified this way was in close agreement with that obtained by experimental methods. GenBank sequences also revealed that (dA-dC)n sequences (n > 6) occur every 18 and 21 kb, on average, in mouse and rat genomes. In order to know if the positions of these repeat sequences were conserved, GenBank DNA sequences containing (dA-dC)n repetitive sequences were compared in several mammalian genomes. In more closely related species such as man/chimp, four out of four comparisons revealed conservation of position. For mouse/rat, five out of seven comparisons revealed conservation of position. There does not appear to be extensive conservation of (dA-dC)n repeat position in evolutionarily distant species, such as rodent/human. Only 6 out of 20 sites where homology could be established were conserved. The few observed sequences of conservation of (dA-dC)n positions, however, indicate that some (dA-dC)n sequences are of very ancient origin.

In 1992 Beckmann and Weber reported a survey of human and rat microsatellites. They summarized GenBank searches for all possible human and rat microsatellites ranging from mono- to tetra-nucleotide repeats. For humans, (A)n·(T)n sequences were most abundant, followed by (AC)n·(TG)n, (A2-3N)n·(T2-3N)n (where N is T, G, or C), (AG)n·(TC)n and (AT)n·(TA)n sequences.

Given the low genomic frequency of CpG di-nucleotides, sequences with the CCG motif were suprisingly abundant. For rat, the most abundant repeat is (AC)n·(TG)n, followed in order by (AG)n·(TC)n, (A2-3N)n·(T2-3N)n, and (A)n·(T)n sequences. The great majority (≥ 80%) of the (A)n·(T)n and (A2-3N)n·(T2-3N)n repeats were associated with *Alu* elements, but only a small fraction of the (AC)n.(TG)n repeats were adjacent to *Alu* elements. Almost all of the microsatellites that were adjacent to *Alu* elements were found immediately 3' to them in the positions normally occupied by the A-rich tail. No microsatellites were found adjacent to the 5' end of *Alu* elements.

### 1.2.1. The Polymorphism of STRs

#### 1.2.2.1. The Length Polymorphism of STRs

Although (AC)n repeat sequences and their variations were found in eukaryotic genomes in the early '80s by Meisfeld et al. (1981) and Hamada et al. (1982b), it was not until the end of the '80s that the length polymorphism of these sequences were independently recognised by Weber and May (1989), Litt and Luty (1989) and Tautz (1989) to represent a wealth of genetic variation applicable to genetic studies. Their observations of the length polymorphism of these sequences, although surprising, greatly accelerated genetic mapping and other studies relating to DNA polymorphisms.

Weber and May (1989) investigated (dA-dC)n·(dT-dG)n repeats by searching the database through the BIO-NET computing network. They screened for the presence of (AC)n or (GT)n sequences with n ≥ 6. To confirm and extend the results from literature sequences, 10 (AC)n repeats were chosen for direct experimental examination. 5 sequences were taken from GenBank and 5 were from human genomic DNA clones which showed (AC)n positive by hybridization. DNA sequences within and immediately flanking the (AC)n repeat sequences were amplified and labelled using the polymerase chain reaction, and were then electrophoresed on standard denaturing polyacrylamide sequencing gels. The IGF1 (AC)n repeat sequences were amplified in seven unrelated individuals. Two of them were homozygous and the remainder were heterozygous. All 10 (AC)n repeats exhibited polymorphism in the size of the amplified fragments. Alleles always differed in size by multiples of 2 bases, a result consistent with the concept that the number of tandem dinucleotide repeats was the variable factor. These 10 polymorphic markers were moderately to highly informative. To determine whether the (AC)n repeat markers exhibited Mendelian codominant inheritance, DNA samples from four 3-generation families were amplified using primers of 4 (AC)n repeat markers. All markers showed Mendelian behavior for all families. From these results Weber and May indicated that 1) there are a very large number of (AC)n repeats in the human genome; this class of polymorphisms is likely to find application in the study of many genetic disease genes and should permit substantial

improvement in the resolution of the human genetic map; 2) the use of the polymerase chain reaction to detect length polymorphisms offers improved sensitivity and speed compared with standard blotting and hybridization.

In the same issue of *American Journal of Human Genetics* in 1989 in which Weber and May reported the investigation of length polymorphisms of (AC)n repeat sequences in the human genome, Litt and Luty (1989) also reported the length polymorphism of the (AC)n repeat sequence in the cardiac muscle actin gene. Because minisatellite sequences with larger repeat elements often display extensive length polymorphism as variable numbers of tandem repeats (VNTR), Litt and Luty assumed (AC)n repeats (microsatellites) might also be polymorphic. To test this idea, an (AC)n microsatellite was chosen from the fourth intron of the human cardiac muscle actin gene. The primers flanking the (AC)n repeat were synthesized. PCR amplification of the microsatellite sequence was performed in a portion of Utah family K1329. Aliquots of the amplified samples were run on a sequencing gel. The autoradiogram showed each of the major bands represented an allele. The Mendelian codominant inheritance of five alleles segregating in this family was observed. From a study of 37 unrelated Caucasian individuals, 12 alleles were detected. Using allele frequencies for calculating of the PIC (polymorphic information content, Botstein 1980), the PIC of this (AC)n marker was 0.86. Therefore Litt and Luty indicated that the high information content and widespread distribution of this type of length polymorphism make microsatellite loci very useful for linkage studies in humans.

Tautz also reported his investigation of length polymorphism of microsatellites in the same year (Tautz, 1989). He investigated di- and tri-nucleotide repeats from *Drosophila melanogaster*, *Globicephala malaena* (a long-finned pilot whale) and human. The (CAG)n repeat sequence was within the coding region of the *Notch* gene of *D. melanogaster*, which was interspersed with CAA triplets (both triplets code for glutamine). The (AG)n·(TC)n repeat sequence was from a randomly isolated whale locus. The two (AC)n·(GT)n repeat sequences were from the intergenic region of the human δ and β-globins and the human cardiac muscle actin gene. The flanking sequences were used to design primers and PCR was performed with end-labelled primers in a standard PCR reaction. PCR products were resolved on a 6% polyacrylamide sequencing gel. Seven alleles were detected for the (CAG)n repeat in 11 lines of *D. melanogaster*. Three alleles were detected for the (TC)n repeat in 11 indviduals of one single population of whales. 3 and 7 alleles were detected for two (AC)n repeats respectively, by genotyping of a human family and 3 unrelated individuals. The family study showed the allele segregation followed a Mendelian inheritance pattern. In his conclusion, Tautz indicated that simple sequence length polymorphisms are effectively 'single locus' probes, can be exploited for genome mapping and linkage studies and can often provide a solution to problems requiring the exploitation of DNA polymorphisms.

Since the length polymorphism of simple tandem repeats (STRs) was observed in 1989, more and more STRs have been found by searching the DNA sequence

databases or by their isolation from mammalian genomes (especially the human, mouse and rat genomes). The exact number of STRs found in eukaryotic genomes is large but not known. The human STR number from the Weissenbach (1993) report alone is 2,794 (AC)n repeat markers. If other reports are included [Engelstein et al., 1993 (chromosome 1); Mills et al., 1992 (chromosome 4); Kwiatkowski et al., 1992 and Wilkie et al., 1992 (chromosome 9); Decker et al., 1992 (chromosome 10); Litt et al., 1993 (chromosome 11); Dawson et al., 1993 (chromosome 12); Petrukhin et al., 1993 (chromosome 13); Beckmann et al., 1993 (chromosome 15), Shen et al., (submitted)(chromosome 16); Strab et al., 1993 (chromosome 18); Hazan et al., 1992 (chromosome 20), McInnis et al., 1993 (chromosome 21); Buetow et al., 1993 (Chromosome 22) and so on], the approximate number of STRs characterised in the human genome has already reached at least 3,000. This number represents a small proportion of the STRs which will eventually be characterised.

It is important to stress the significant role of the development of the PCR technique (Saiki et al., 1985), with which the length polymorphism of STRs was discovered, and with which these length polymorphisms can be applied as very useful genetic markers for many purposes.

1.2.2.2. The Informativeness of STRs

As the human genome initiative progresses, more and more STRs [especially (AC)n repeat sequences] have become available. The ability to predict the genetic informativeness and the utility of a specific repeat on the basis of sequence alone is important to improve the efficiency of new marker development. Weber (1990a) reported on the informativeness of human (AC)n·(TG)n polymorphisms. He examined over 100 human (AC)n·(TG)n sequences with six or more consecutive (AC) or (TG) repeats from a computer search of GenBank and from sequencing M13 clones selected by stringent hybridization to poly(AC)·poly(TG). By establishing rules for categorizing (AC)n·(TG)n sequences, these were divided into three categories: 1) perfect repeat sequences were defined as AC repeats without interruption and without adjacent repeats of another sequence; 2) imperfect repeat sequences were defined as AC repeats with one or more interruptions in the run of the repeat sequence; 3) compound repeat sequences were defined as AC repeats with an adjacent repeat of a different sequence. The results showed the perfect repeat sequences to be predominant (64%), imperfect repeat sequences next most frequent (25%) and compound repeat sequences relatively infrequent (11%). For the perfect category of repeat sequences, the informativeness for sequences with 10 or fewer repeats was very low or zero, with 11-15 repeats was quite variable and with 16 or more repeats was moderate to high. Imperfections in the repeat sequences tended to reduce the informativeness of the resulting polymorphisms relative to that which

would be expected for the same total number of perfect repeats. The best method for predicting the polymorphism of imperfect repeat sequences appears to be to take the longest run of uninterrupted repeats as the length of the repeat sequence. PIC values were determined for only three compound repeat sequences. The values were all relatively high (because the perfect repeat number in the compound sequence was not shown in the report, it was hard to know whether the high PIC value resulted from a long perfect or compound repeat.). From these results, it is clearly demonstrated that the informativeness of (AC)n·(TG)n repeat sequences generally increases as the average number of repeats increases.

Weber (1990a) also used the results from the hybridization and distribution of cloned sequences to estimate the number of highly informative (AC)n·(TG)n sequences in the human genome. There are about 12,000 (AC)n·(TG)n repeat sequences which should be found in the human genome leading to polymorphisms with PIC values of $\geq 0.50$ and about 7,000 with PIC $\geq 0.70$. The total genetic length for the human genome is 3300 cM (White and Lalouel, 1988). These numbers of (AC)n·(TG)n repeat markers alone would produce a human genetic map with an average resolution of approximately 0.3-0.5 cM., assuming random distribution throughout the genome.

Because the simple tandem repeat sequences other than (AC)·(TG)n also exhibit length polymorphisms, it is likely that informativeness of polymorphisms based on

these other sequences will show qualitatively similar dependence on repeat sequence length and type as the (AC)·(TG)n repeat polymorphisms. Zuliani and Hobbs (1990) and Edwards et al. (1991) reported tri- and tetra-nucleotide repeat polymorphisms. The heterozygosities of (TTA)n, (AAT)n, (AGC)n, (AATG)n, (ACAG)n and (AGAT)n repeat sequences ranged from 0.34 to 0.89. The informativeness of these repeat sequences also showed an increase as average repeat number increased.

1.2.3. The Possible Functions of the STRs

Since simple tandem repeat sequences are abundant, interspersed sequences in eukaryotic genomes ranging from yeast to human, this suggests that these ubiquitous elements may have biologically functional significance. The following are some possible functions postulated for these sequences.

A) Gene Regulation and DNA Replication

Hamada and Kakunaga (1982a) reported that potential Z-DNA forming sequences [e.g. (dA-dC)n·(dT-dG)n sequence] are highly repeated in the human genome. They indicated that if Z-DNA occurs naturally then it might have quite different reactivities with molecules such as proteins or carcinogens from right-handed B-DNA. The interconversion of sequences between B and Z forms, under

the influence, for example, of DNA binding proteins or chemical modification, may be important in regulating DNA function.

A possible role for poly(AC)n sequences in transcription was detected by Hamada et al. (1984). They tested the hypothesis that the (AC)n elements can modulate gene expression. The human genomic DNA fragments containing a (dA-dC)n·(dT-dG)n sequence (30-50 bp) or chemically synthesized (dA-dC)n·(dT-dG)n sequence (50-130 bp) were inserted into the pSV2-cat (simian virus 40 enhancer plus) or pA10-cat (enhancer minus) expression vector plasmid. These constructs were transfected into CV-1 cells or HeLa cells, and their transcription was monitored by assaying reporter gene (chloramphenicol acetyltransferase) activity. The results showed that pSV2-cat with the (AC)n element and pA10-cat with the (AC)n element synthesized more chloramphenicol acetyltransferase activity [2-10 times, depending on the location of the (AC)n element] than did parental pSV2-cat and pA10-cat DNA, respectively. Furthermore, the (AC)n element appeared to have characteristics similar to those of viral enhancers: 1) the (AC)n element enhanced transcription from a distance, 2) its closer location to the promoter was more effective, and 3) its orientation was not crucial. These results suggested that the (AC)n element may influence the expression of cellular genes.

In order to know whether any positions of poly(AC)n were conserved in mammalian genes, Braaten et al. (1988) analyzed the ribosomal DNA of mouse, rat

and human by hybridizing poly(AC)n probe. The poly(AC)n sequences were found in comparable locations upstream and downstream of the transcription units among the ribosomal DNAs of these species. These authors suggested that poly(AC)n sequences occurred in specialised regions (e.g. introns) that served some function involving gene regulation.

Because the function of simple tandem repeat sequences is still unknown, it has only been in those instances where their physical location within a gene is associated with a disease phenotype that some insight into their function has been gained. Recently Richards et al. (1993) reported that the fragile X syndrome unstable element, p(CCG)n, and other simple tandem repeat sequences are binding sites for specific nuclear proteins. Four dinucleotide and ten trinucleotide repeats were analysed using bandshift assays. All the repeats tested were capable of being bound by nuclear proteins which have been termed simple tandem repeat-binding proteins (STR-BPs). For fragile X syndrome, the heritable unstable sequence p(CCG)n is located in the 5' untranslated region of the gene (*FMR1*) in which it is transcribed. When the amplification of the p(CCG)n repeat occurs, transcription of the *FMR1* gene is terminated and flanking sequences hypermethylated. The bandshift assays showed that the fragile X unstable element p(CCG)n in either its double or single strand form binds a number of nuclear proteins. The single stranded binding proteins are of particular interest in relation to the mechanism of mutation and expression of the fragile site, since it has long been thought that induction of the

fragile site could be a manifestation of under-replication of fragile site DNA sequences (Sutherland, 1988). The finding of multiple nuclear proteins for simple tandem repeats further supports the hypothesis of a role for these sequences. Richards et al. indicated that instead of a role in the level of transcription *per se* that the repeats might normally be involved in coupling *FMR1* transcription to some other cellular process, such as replication. The existence of strand specific single-stranded binding proteins may be indicative of such a role. The finding that methylation specifically interferes with the binding of CCG.BP1 and allows the binding of a different protein (possibly the non-sequence specific MeCP1) suggests a molecular pathway which can account for the coupling of methylation to lack of *FMR1* expression. Amplification of the repeat may hinder transcription (since the repeat sequences are able to form unusual structures) and associate with at least partial methylation of the repeat. The partially methylated repeat then becomes a lower affinity binding site for CCG.BP1 and a higher affinity site for the methyl-p(CCG)n binding protein. If this protein is MeCP1 then its binding would further suppress *FMR1* transcription, since MeCP1 is a known repressor of transcription. Richards et al. further indicated CCG.BP1 need not take an active role in transcription but merely be a facilitator of the process, perhaps by protecting the p(CCG)n repeat from methylation. All of the 14 STR sequences tested in this bandshift assay were found to be protein binding sites, and two repeats, p(CCG)n and p(AGC)n (myotonic dystrophy), are located in the untranslated regions of the

genes (*FMR1* and *DM1*, respectively). Their ability to act as specific sites for DNA binding proteins implies they take a role in some cellular function such as gene regulation and/or DNA replication.

B) DNA Recombination and Gene Conversion

Slightom et al. (1980) investigated the nucleotide sequences of the $^{G}\gamma$- and $^{A}\gamma$-globin genes from two chromosomes of one individual. They found two simple repeat sequences of $(AC)_{19}$ and $(AC)_{13}$ located in the $^{G}\gamma$ gene and the $^{A}\gamma$ gene of one chromosome (A), a $(GT)_{17}$ in the $^{A}\gamma$ gene of other chromosome (B). The repeats on chromosome (A) can be formally related to each other by a deletion resulting from an unequal exchange between the two (AC)n repeats with the loss of 20 bases. Because the similarities and differences between the allelic $^{A}\gamma$ genes and the nonallelic $^{A}\gamma$ and $^{G}\gamma$ genes were reversed on the two sides of this region, Slightom et al. suggested that this simple sequence region was a hot spot for recombination which led to gene conversion.

Pardue et al. 1987 reported the results from hybridization of (dA-dC)n·(dT-dG)n to the polytene chromosomes of *Drosophia* species. They found the complete lack of hybridization of AC/TG sequences to β-heterochromatin. This region had the property of not undergoing meiotic recombination. It raised the possibility that the lack of meiotic recombination may be correlated with the lack of AC/TG sequences. In contrast to the other species, there was a AC/TG hybridization over the

chromosome of *D. virilis*. In this case the presence of AC/TG correlated with the ability to undergo meiotic recombination. The recombination in *D. virilis* is 2.8-fold more frequent than in *D. melanogaster*, and the AC/TG content was found higher in *D. virilis*.

C) Chromatin Structure

The (AC)n·(TG)n were distributed over the euchromatin regions, but were not detectable in the heterochromatin regions in *Drosophia* species (Pardue et al. 1987) or in humans (Stallings et al. 1991). This suggested that (AC)n·(TG)n sequences may be, at least in part, responsible for the regional attributes. Because very large regions of the chromosomes are involved, the (AC)n·(TG)n sequences may act by affecting chromatin structure. The (AC)n·(TG)n sequences have been shown to be capable of adopting a Z-DNA conformation within negatively supercoiled plasmids (Nordheim and Rich 1983) and they are also able to adopt other conformations. They were found to be not in Z-DNA conformation within cells and appeared to quantitatively adopt a distinctive B-DNA conformation on the nucleosomal surface (Gross et al. 1985).

These functions of simple tandem repeat sequences are of interest but they have not yet been directly proved. The heritable unstable sequences [e.g. p(CCG)n and p(AGC)n ] have been found in the gene for the fragile X syndrome (Kremer et al., 1991), myotonic dystrophy (Brook et al., 1992), spinal and bulbar muscular

atrophy (Kennedy disease, La Spada et al., 1991), Huntington disease (MacDonald et al., 1993) and spinocerebellar ataxia type 1 (SCA1, Orr et al., 1993). Mutations of some STRs have been found in colorectal cancer cells (Peltomaki et al., 1993; Thibodeau et al., 1993). It is expected that some functions of STRs will be revealed in these instances where the STR instability is associated with a disease phenotype.

## 1.3. THE APPLICATIONS OF THE STR POLYMORPHISMS

### 1.3.1. The Construction of the Genetic Linkage Maps of Genomes

Since the simple tandem repeats were observed to show Mendelian codominant inheritance and to be highly polymorphic in eukaryotic genomes in 1989, they have been used as genetic markers for the construction of high-resolution genetic maps and for improvement in the resolution of existing genetic maps. In 1991 Weber et al. published the first map of human chromosome 5 based on 13 simple tandem repeat markers. A panel of human-mouse somatic cell hybrids was used to physically map STR markers to cytogenetic regions of human chromosome 5. The genetic order of STR markers was determined by linkage analysis. These 13 STR markers covered about 90% of chromosome 5.

With the initiative of a global Human Genome Project established, construction of the genetic linkage maps of human chromosomes aimed to develop maps containing highly polymorphic markers spaced 2-5 cM apart in the first five years

(the official 5-year plan was set to begin on October 1, 1990 and to end on September 30, 1995) (Jordan, 1992). In three years, about twenty genetic linkage maps of human chromosomes based on STR markers were published [ Engelstein et al., 1993 (chromosome 1, 38 STRs); Mills et al., 1992 (chromosome 4, 16 STRs); Weber et al., 1991 (chromosome 5, 13 STRs); Wilkie et al., 1993 (chromosome 6, 11 STRs); Tomfohrde et al., 1992 (chromosome 8, 22 STRs); Emi et al., 1993 (chromosome 8p, 14 STRs); Kwiatkowski et al., 1992 (chromosome 9q, 14 STRs); Wilkie et al., 1992 (chromosome 9, 10 STRs); Deker et al., 1992 (chromosome 10, 10 STRs); Litt et al., 1993 (chromosome 11, 25 STRs); Dawson et al., 1993 (chromosome 12, 22 STRs); Petrukhin et al., 1993 (chromosome 13, 21 STRs); Bowcock et al., 1993 (chromosome 13q, 12 STRs); Wang and Weber, 1992 (chromosome 14, 9 STRs); Beckmann et al. (chromosome 15, 55 STRs); Shen et al. submitted (chromosome 16, 78 STRs); Straub et al., 1993 (chromosome 18, 14 STRs); Hazan et al., 1992 (chromosome 20, 26 STRs); McInnis et al., 1993 (chromosome 21, 43 STRs); Buetow et al., 1993 (chromosome 22, 24 STRs)]. Although the resolution of most of these maps has not reached the first 5-year goal of the Human Genome Project (the maps of chromosomes 15, 16 and 21 have an average interlocus resolution of 2.0, 3.2 and 2.5 cM, respectively), the highly informative markers on these maps are useful at the present time and will contribute to the construction of a high-resolution genetic map of the human genome in the five year period. In November, 1992 Weissenbach et al. published a genetic linkage map

consisting of 813 (AC)n repeat markers covering every chromosome (except the Y chromosome). This map was a milestone in human genome mapping. More recently, Weissenbach (1993) reported that a total of 2794 (AC)n repeat markers randomly isolated in his laboratory have been assigned to 23 human chromosomes. It is expected that a map containing about 3,000 (AC)n repeat markers will be soon published.

The construction of genetic linkage maps based on STR markers from other organisms have also been underway. The mouse and rat genetic linkage maps, which contained 317 and 174 STR markers respectively, have been constructed (Dietrich et al., 1992; Serikawa et al., 1992). Simple tandem repeat sequences have also been isolated from livestock genomes. There are more than 400 STR markers isolated from cattle, around 200 in pigs, less than 100 in sheep and chicken (Hetzel, 1993). The primary linkage maps are close to completion for all the major livestock species (Hetzel, 1993). (AC)n repeat sequences have been also isolated from the dog genome (Ostrander et al., 1993) and the zebrafish genome (Goff et al., 1992) for use in the construction of genetic maps for these species.

The more simple tandem repeat sequences are isolated, the higher will be the resolution of genetic maps so constructed. The initial goals of the Human Genome Project and other organism (e.g. animals and plants) genome mapping projects will be successfully achieved in the near future.

### 1.3.2. The Genetic Localization of the Disease Genes

The initial step in gene localization for familial disorders for which the biochemical basis is unknown is a linkage study. It was because so few genetic markers (e.g. blood groups, enzyme and protein polymorphisms) were available before 1980 that the chromosomal assignment and regional localization of disease genes by linkage study was relatively inefficient by present standards. After Botstein et al. (1980) published the potential use of RFLPs for linkage, RFLPs were soon utilised to localize the genes of several genetic diseases, including Huntington disease (Gusella et al., 1983), Duchenne muscular dystrophy (Davies et al., 1983), cystic fibrosis (Tsui et al., 1985) and adult polycystic kidney disease (Reeders et al., 1985). Although RFLPs have allowed significant progress to be made in human genetics, their limitations did not allow them to be ideal genetic markers for linkage studies. With the advent of STR polymorphisms and the PCR technique, the number of disease genes assigned to chromosomes and given refined localization by linkage study has dramatically increased. Here, only a few diseases are taken as examples.

Alkaptonuria, the first human disorder recognized by Garrod in 1908 as an inborn error of metabolism, is a rare autosomal recessive disorder characterized by the inability to metabolize homogentisic acid. Rare recessive diseases in the offspring of consanguineous parents usually arise because the affected individuals

have two defective gene copies from the common ancestor and therefore flanking genetic markers near the responsible locus should also be homozygous. Lander and Botstein (1987) coined the term "homozygosity mapping" to describe a strategy for chromosomal localization of recessive disease genes. The coinheritance of alkaptonuria and neonatal severe hyperparathyroidism (NSHPT), which is thought to be the recessive form of familial hypocalciuric hypercalcemia (FHH) has been hypothesized to reflect close linkage of the loci responsible for both disorders (Pollak et al., 1993). FHH is an autosomal dominant condition caused by mutation in a gene on chromosome 3q21-q24 (Pollak et al., 1993). The coinheritance of alkaptonuria and sucrase-isomaltase (SI) deficiency have also been reported and the gene responsible for SI has been localized to chromosome 3q26 (Pollak et al., 1993). Based on these reports, Pollak et al. (1993) used 15 simple tandem repeat markers in 3q2 region to map the gene responsible for alkaptonuria. These markers were genotyped through two families with consanguineous parents and four affected children. Each of four affected individuals were homozygous for all six loci between *D3S1303* and *D3S1292*. The homozygous haplotypes were used as a single locus for linkage calculation. The lod score in the two families is 3.92 at $\theta=0$. Using homozygosity mapping, the alkaptonuria gene was localized in a 16 cM region on chromosome 3q2 region.

Spinocerebellar ataxia 1 (SCA1) is one form of autosomal dominant cerebellar ataxia (ADCA), which is a group of hereditary neurodegenerative disorders.

Thirteen years ago, Morton et al (1980) localized SCA1 on chromosome 6p based on linkage to the HLA loci. In 1991, SCA1 was found to be very closely linked to the (AC)n repeat marker *D6S89* and localized to 6p22-p23 (Zoghbi et al., 1991; Ranum et al., 1991). The genetic distance between SCA1 and *D6S89* was still too great (8 cM) to permit positional cloning of this gene. Five of the genetically mapped markers reported by Weissenbach et al. (1992) are very closely linked to *D6S89* in a 6 cM interval. Jodice et al. (1993) used these 5 markers for the genotyping of 8 large Italian SCA1 kindreds. The results of linkage analysis showed that the interval containing the SCA1 gene can be reduced to 3 cM flanked by *D6S247* and *D6S89*. In efforts to clone the SCA1 gene, a USA team (Orr et al., 1993) developed a YAC contig, which extended from *D6247* to *D6S89* and spanned the entire SCA1 candidate region. During the positional cloning, cosmid clones were screened with trinucleotide repeats because anticipation was observed in some SCA1 families. Anticipation is a characteristic associated with diseases caused by dynamic mutation of simple tandem repeat sequences. 23 clones were shown to be positive for the CAG repeat probe. To test the genetic stability of this repeat in SCA1, Southern blotting analysis was used to examine families with SCA1. The CAG repeat was found to be unstable and expanded in the individuals with SCA1. This was the fifth genetic disease found to show dynamic mutation.

For localization of genes of major effect contributing to multifactorial diseases, genetic markers must be highly polymorphic because families with multiple living

affected relatives and parents are rare, and difficult and expensive to collect. To date highly polymorphic STR markers have already contributed to the localization of new genes for familial Alzheimer's disease (Schellenberg et al., 1992) and diabetes (Froguel et al., 1992). A locus predisposing to human colorectal cancer has been localized on chromosome 2 by a systematic search through the whole genome with 345 (AC)n repeat markers (Peltomaki et al., 1993). Linkage analysis of colorectal cancer to microsatellite (AC)n repeats on chromosome 2 was demonstrated in two large kindreds. The pairwise lod scores for linkage to marker *D2S123* in these two kindreds were 6.39 and 1.45 at $\theta = 0$, and multipoint linkage with flanking markers resulted in lod scores of 6.47 and 6.01. More recently, a second locus predisposing to hereditary non-polyposis colon cancer has been found to be tightly linked to an (AC)n repeat marker *D3S1029* on chromosome 3p (Lindblom et al., 1993).

From these examples, it has been shown that highly polymorphic STR markers have facilitated and speeded gene localization and gene cloning. It will be seen that more and more genes responsible for genetic disorders, including monogenic and polygenic diseases, will be localized by linkage analysis using STR markers.

### 1.3.3. Other Applications

#### A) Detection of Loss of Heterozygosity (LOH) in Cancers

Loss of heterozygosity (LOH) of a specific chromosomal region is generally considered to be involved in genesis and progression of cancer through a mechanism of inactivation of tumor suppressor genes which are located in the region. The highly polymorphic STR markers have been used to detect LOH in several cancers. The following are examples of familial breast and ovarian cancer and colorectal cancer.

A gene associated with susceptibility to familial breast and ovarian cancer was recently localized on chromosome 17q12-21. Smith et al. (1992) reported loss of heterozygosity in tumours of breast and ovarian cancer. They typed four (AC)n repeat markers in this region by PCR using lymphocyte DNA and DNA extracted from tumour tissues. Of the two breast tumours and eleven ovarian tumours examined in four affected families, one breast tumour and eight ovarian tumours showed allele losses for these markers. The result suggested that the putative breast-ovarian cancer gene is a tumour suppressor gene.

Recently, Aaltonen et al. (1993) investigated the molecular features of familial colon cancers and sporadic colon cancers by detecting LOH using (AC)n repeats. Neither the familial colorectal cancer nor sporadic cancers were found to show loss

of heterozygosity for chromosome 2 markers. However, they found that 1) the incidence of mutations in *KRAS, P53* and *APC* was similar in two groups of tumours, and 2) most of the familial cancers had widespread alterations in (AC)n repeat sequences. From these results, Aaltonen et al. suggested that numerous replication errors had occurred during tumor development, and there was a mechanism for familial tumorigenesis different from that mediated by classic tumour suppressor genes.

These examples have demonstrated that simple tandem repeats are a useful tool for localizing cancer suppressor genes and revealing other possible genetic mechanisms of cancer development.

B) Prenatal Diagnosis

For the molecular prenatal diagnosis of genetic diseases and determination of carrier status, STR markers have demonstrated the advantages of being more rapid and more informative than the RFLP markers. For example, the average level of heterozygosity at eight RFLP sites in phenylketonuria (PKU) families was about 36% in the Chinese population. Goltsov et al. (1993) reported that they identified a novel (AGAT)n repeat, within the PAH gene, with 9 alleles, which had an average level of heterozygosity of about 75% in the Chinese population and 80% in the European population. This marker allowed rapid and efficient prenatal diagnosis and carrier screening for phenylketonuria.

To improve carrier detection and prenatal diagnosis for Duchenne and Becker muscular dystrophy families, Clemens et al. (1991) identified four (AC)n repeat sequences in introns 44, 45, 49 and 50 of the dystrophin gene. These markers aided in the identification of deletion mutation, exclusion of maternal cell contamination of CVS, confirmation of paternity and mapping of genetic recombination.

The (AC)n repeat markers have also been used to detect the parental origin of the additional chromosome 21 in Down syndrome (Petersen et al., 1991), to demonstrate paternal nondisjunction in trisomy 21 (Petersen et al., 1993) and to demonstrate isodisomy in Prader Willi syndrome and Angelman syndrome (Trent, 1993).

C) Evaluation of Linkage Disequilibrium

With the analysis of haplotypes of fragile X chromosomes from unrelated fragile X individuals in Australia and USA, two (AC)n repeat markers flanking the FMR1 gene were found in linkage disequilibrium with each other and with the (CCG)n repeat (Richards et al., 1992). This finding showed evidence of founder chromosomes in the fragile X syndrome.

Three (AC)n repeat markers in the region of the Batten disease gene (CLN3) were found to have at least one allele in strong linkage disequilibrium with the disease locus (Mitchison et al., (1993). This led to a refined localization of CLN3 on

chromosome 16. Analysis of haplotypes at these loci suggested that the majority of CLN3 chromosomes have arisen from a single mutational event.

D) Population Studies

In 1992 Edwards et al. reported genetic variation in different populations. They analysed 5 STR markers (tri- and tetranucleotide repeats) in U.S. Black, White, Mexican-American and Asian populations. The relative allele frequencies exhibited differences between populations with unimodal, bimodal and complex distributions.

Wall et al. (1993) also reported the investigation of the allele frequency distributions of three STR markers in five populations: North European, Crypriot, Pakistani, Gujarati and Vietnamese. They found variation of STRs within and between populations. Their results indicated that in general STR markers are polymorphic in widely differing ethnic groups, and some may be of value in distinguishing ethnic sub-groups by analysis of allele frequency.

E) Forensic Medicine

For forensic medicine, recently STRs (as well as VNTRs) have also been applied in paternal testing and zygosity testing, and in identification of criminal suspects (Trent, 1993).

## 1.4. THE PUBLISHED GENETIC MAPS OF HUMAN CHROMOSOME 16

In 1987 Donis-Keller et al. published the first genetic linkage map of the human genome, which contained every chromosome except the Y chromosome. The chromosome 16 genetic map consisted of 42 markers (one was a VNTR marker, 3'HVR). This first map of chromosome 16 covered almost the whole chromosome from HBA1 on 16p to *D16S44* (CRI.089) on 16q. The genetic length of the sex-averaged map was 195 cM. The lengths on the male and female maps were 164 and 237 cM, respectively. There were 6 gaps larger than 10 cM, with one gap larger than 30 cM on 16q.

In 1990 Keith et al. published a genetic linkage map of chromosome 16 which contained a total of 46 markers (45 RFLPs, 1 VNTR), 42 of these were the same as markers on Donis-Keller's map. In comparison with the Donis-Keller's map, the length of the sex-averaged map was reduced to 149 cM and 3 gaps were larger than 10 cM. The lengths of the male and female maps were 115 and 193 cM, respectively.

Julier et al. also reported a genetic map of chromosome 16 in 1990. The map contained 24 markers (20 RFLPs, 4 VNTR), 21 of them were new RFLP and VNTR markers. The genetic lengths on the male and female maps were 186 and 226 cM respectively, and no sex-averaged genetic length was reported.

NIH/CEPH Collaborative Group (1992) published the comprehensive genetic linkage map of the human genome, which consisted of 23 genetic maps of human chromosomes. The chromosome 16 map (contributed by Kozman et al.) was based on 67 markers (61 RFLPs, 2 VNTRs,4 STRs) from 50 loci. These loci were placed in 50 cytogenetic intervals along this chromosome and genotyped on the CEPH reference families. The sex-average map is 162 cM long, 133 cM and 199 cM in male and female maps, respectively.

Kozman et al. (1993) published a genetic linkage map of chromosome 16 based on NIH/CEPH map (all markers were same). The genetic lengths were slightly different: the sex-averaged map was 165 cM, and the male and female maps were 133 and 202 cM, respectively.

These genetic linkage maps of chromosome 16 were all mainly based on RFLP markers. Thus, their utility was limited for assignment and localization of disease genes. In 1992 Weissenbach et al. published the second-generation linkage map of the human genome which was based on 813 highly polymorphic STR markers [(AC)n repeats]. These markers were assigned to 22 autosomes and the X chromosome. The distance spanned about 90% of the estimated length of the human genome. The chromosome 16 map contained 24 (AC)n repeat markers, which did not cover the whole chromosome, with a genetic length of 116 cM. The physical locations of these markers on chromosome 16 were not known.

In order to more efficiently localize disease genes on chromosome 16 and for other purposes involving DNA polymorphisms, there was a need for construction of a high-resolution genetic linkage map of chromosome 16 based on STR markers.

## 1.5. THE APPROACHES USED FOR CONSTRUCTION OF GENETIC MAPS BASED ON STR MARKERS

To construct a genetic linkage map based on STR markers isolation of STR markers is the essential step, followed by characterization, physical mapping, genotyping of reference families and linkage analysis.

For isolation of STR sequences, the general approach used was random isolation (Weber et al., 1991; Thompson et al., 1992; Weissenbach et al., 1992; Petrukhin et al., 1993). The chromosome specific libraries or human genomic libraries were screened with (AC)n or other di, tri- and tetranucleotide sequences. The positive clones were sequenced and primers designed for PCR.

PCR techniques are used for the characterization, physical mapping and genotyping of STR markers. The heterozygosity is determined by allele frequencies in unrelated individuals (Weber, 1990a). The locations of STR markers are assigned by physical mapping using somatic cell hybrids (Thompson et al., 1992). CEPH (Centre d'Etude du Polymorphisme Humain) reference families (Dausset et al., 1990) are used for genotyping.

The linkage analysis is performed using computer programs, for example, CRI-MAP (Lander and Green, 1987; Green et al., 1990), LINKAGE (Lathrop and Lalouel, 1984) and MAPMAKER (Lander et al., 1987). The map distances and marker orders are determined by likelihood analysis.

To date, using similar approaches to those described here, twenty one genetic linkage maps of individual human chromosomes have been constructed (see section 1.3.1. of this chapter).

## 1.6. AIMS OF THE PROJECT

The aim of the project was to construct a high-resolution genetic map of human chromosome 16 with STRs, and to construct an STR physical map based on a high resolution cytogenetic-based physical map of chromosome 16. The average resolution aimed for was 2-5 cM between markers which was in accord with one of the first five year goals of the International Human Genome Project.

During development of these maps, relevant STRs were identified and used to refine localization of the gene (*MEF*) responsible for familial Mediterranean fever (FMF) at chromosome 16p13.3.

To achieve the aim of construction of an STR-based genetic and physical maps of chromosome 16, 1) isolation of STR markers was carried out using random and targetted approaches; 2) characterization of STR markers was determined by allele

frequencies and heterozygosities in the Caucasian population; 3) STR markers were physically mapped on a high-resolution cytogenetic map (Callen et., 1992) using a human-mouse somatic cell hybrid panel (Callen et al., 1992); 4) genotyping of STR markers was performed using CEPH reference families (Dausset et al., 1990); and 5) linkage analysis was carried out to construct a genetic map of human chromosome 16 using the computer program CRI-MAP (Lander and Green, 1987; Green et al., 1990).

To achieve the aim of refining localization of the *MEF* gene, genetically and physically mapped STR markers in chromosome band 16p13.3 were selected for genotyping FMF families. The genotype data were analysed using recombinant haplotypes, homozygosity mapping and computer program LINKAGE (Version 5.1) (Lathrop et al., 1984).

## 1.7. SUMMARY

Simple tandem repeat sequences are widely-dispersed elements in eukaryotic genomes. There are about 50,000 (AC)n repeat sequences in the human genome. They exhibit length polymorphisms, Mendelian codominant inheritance (except where null alleles are detected), are generally highly informative, and are easy to use with the PCR technique. STR sequences represent an abundant source of DNA polymorphisms.

The biological functions of the STR sequences are still unknown. Some STR sequences are associated with genetic diseases, for example, the heritable unstable trinucleotide repeats [e.g. p(CCG)n, p(AGC)n] in fragile X syndrome, myotonic dystrophy, Kennedy disease, Huntington disease and spinal cerebellar ataxia (SCA1). This, together with the finding of STR instability in cancers, will help to reveal some of these functions in the near future.

The characteristics of STR sequences make them ideal genetic markers for construction of high-resolution maps of the human genome, for efficient localization of disease genes, for detection of genome anomalies in cancers, for improvement of prenatal diagnosis and carrier determination of genetic diseases and for population studies.

The project of construction of a high-resolution cytogenetic based physical and genetic map of human chromosome 16 based on STR markers was undertaken. The mapped STRs in 16p13.3 region were applied to refine localization of the *MEF* gene.

# CHAPTER 2

## MATERIALS AND METHODS

## 2.1. INTRODUCTION

This project used molecular genetic technology to isolate and characterize highly polymorphic simple tandem repeat markers (STRs) for the construction of a high-resolution genetic linkage map of human chromosome 16. Most of the methods were well established and used routinely in the laboratory of molecular genetics at the Women's and Children's Hospital (Adelaide, Australia). In this chapter the materials and methods used throughout the whole project will be described. The technologies and approaches involved in isolation, sequencing, characterization, physical mapping, genotyping of STR markers and linkage analysis will be described in the corresponding chapters.

The work on some STR markers was in collaboration with colleagues (indicated in the thesis). For construction of a high-resolution genetic map of human chromosome 16, 32 STR markers from other laboratories were incorporated in the linkage analysis. These markers were physically mapped by the candidate and some of them were characterized and genotyped by the candidate.

## 2.2. ISOLATION OF DNA

### 2.2.1. DNA Isolation of Chromosome 16 Cosmid Library (modification of the method of Maniatis et al. 1989)

The chromosome 16 cosmid library was provided by Dr. R.L. Stallings of Los Alamos National Laboratory in the USA. The concentration of the library was estimated at $1x10^6$ cfu/µl. 2 µl of the glycerol stock was diluted into 5 ml Luria Bertani (LB) medium. 0.5 ml of the 5 ml was diluted into 9ml LB medium (1ml contained about 20,000 cfu). 1ml of the 9.5ml dilution was plated onto each of 9 LB plates containing ampicillin (50 µg/ ml). The plates were incubated at 37°C overnight. 10 ml LB Medium was added onto a plate and left at 4°C for 45 minutes. The colonies were scraped from each plate and were transferred to 500 ml centrifuge tubes. 5 ml more of LB medium was added onto plates for recovering remaining cells and transferred to the same 500 ml centrifuge tube. The cell suspension was spun in a Jouan CR3000 centrifuge at 1,600g at 4°C for 15 minutes. After discarding the supernatant, the cell pellet was resuspended in 6 ml TE (50 mM Tris-HCl, 20 mM EDTA) containing 100mg lysozyme (Boehringer) and left on ice for 10 minutes. 12ml of 0.2 M NaOH / 1% sodium dodecyl sulphate (SDS) was added to the cell suspension, gently mixed and left on ice for another 10 minutes. 7.5 ml of ice cold 3 M sodium acetate (pH 5.2) was added to the suspension, mixed by inverting, and left at room temperature for 30 minutes. The

mixture was spun in a Beckman J2-21 M/E centrifuge with a JA20 rotor at 27,200g for 20 minutes. The supernatant was transferred to a fresh tube. 20 µl of RNase A (10mg/ml) was added to the supernatant and incubated at 37°C with vigorous shaking for 1 hour. For DNA extraction, an equal volume (to DNA solution) of phenol (previously distilled, then saturated with 10 mM Tris HCl) was added and gently mixed with the DNA solution by inverting for 5 minutes. The mixture was then spun at 27,200g for 6 minutes.The upper aqueous phase was transferred to a fresh tube. The above step was repeated twice. An equal volume of chloroform: isoamyl alcohol (24:1) was added and the mixture then spun at 27,200g for 25 minutes. The aqueous phase was divided into two oakridge tubes and 20 mls of ethanol was added. The mixture was left at room temperature for 30 minutes and spun at 27,200g for 25 minutes. The DNA pellet was resuspended in 1.6 ml $H_2O$, 0.4 ml of 5 M sodium chloride and 2 ml of 13% polyethylene glycol (PEG) 6,000, mixed and stored in ice-water for 1 hour. The mixture was spun at 27,200g for 10 minutes. The supernatant was discarded. 8.2 ml of absolute ethanol was added and left at 4°C overnight. After centrifugation, the DNA pellet was washed twice with cold 70% ethanol, air-dried and resuspended in 400 µl TE.

2.2.2. Isolation of Cosmid DNA (modification of the method of Maniatis et al. 1989)

50 ml LB medium containing ampicillin (50 µg/ml) or kanamycin (50 µg/ml) was incubated with a single bacterial colony . The culture was incubated at 37°C overnight with vigorous shaking. The culture was transferred to a 50 ml Beckman tube. The tube was left on ice for 10 minutes and then spun at 1,600g for 15 minutes at 4°C. The supernatant was discarded. The cell pellet was gently resuspended in a solution containing 25 mM Tris-HCl/10 mM EDTA/15% sucrose and 8 mg lysozyme and left on ice for 30 minutes. 8 ml of 0.2 M NaOH/1% SDS was added, mixed gently and left on ice. After 30 minutes, 5 ml of 3 M pH 5.2 sodium acetate was added into the mixture solution which was then left on ice for 10 minutes. The solution was spun in a Beckman centrifuge at 27,200g for 30 minutes.The supernatant was transferred to a 50 ml falconer tube. The DNA extraction was performed twice with an equal volume of phenol and twice with an equal volume of chloroform : isoamyl alcohol (24 : 1). For DNA precipitation, 1/10 volume of 3 M sodium acetate (pH 5.2) and 2 volumes of cold ethanol were added to the tube, inverted several times and left at -70°C for at least 1 hour. The DNA precipitation was spun at 27,200g for 15 minutes. The supernatant was discarded. The DNA pellet was resuspended in 0.5 ml $H_2O$ with 5 µl RNase A (10 mg/ml), placed on a rotating wheel (10 rpm) and incubated at 37°C for 20 minutes. The phenol/chloroform DNA extraction was performed as described above. For DNA

precipitation, 1/10 volume of 3 M sodium acetate and 2 volumes of cold ethanol were added, and left in a -70°C freezer for 1 hour or -20°C overnight. The DNA was washed twice with 70% ethanol, and after desiccation the DNA was dissolved in 200 µl of TE (10 mM Tris-HCl/0.1 mM EDTA).

### 2.2.3. Isolation of Peripheral Lymphocyte DNA (modified method of Wyman and White, 1980)

Blood samples were collected in 10 ml tubes containing EDTA and were allowed to cool to room temperature before being stored at -70°C. For isolating lymphocyte DNA, the frozen blood sample was thawed and transferred to a falconer tube. Cell lysis buffer (0.32 M sucrose/10 mM Tris-HCl/5 mM $MgCl_2$/1% Triton X-100) was added to the tube until the 30 ml mark was reached. After mixing, the tube was left on ice for 30 minutes. The cell suspension was spun in a Jouan centrifuge (4°C) at 2,200g for 15 minutes. The supernatant was aspirated down to 5 ml then cell lysis buffer was added again to the 30 ml mark. Centrifugation was repeated once more. The supernatant was carefully removed, 3.25 ml of Proteinase K buffer, 0.5 ml of 10% SDS and 0.2 ml of Proteinase K (10mg/ml) were added and mixed well with the cell pellet. The tube containing the cell suspension was sealed with parafilm, placed on a rotation wheel (10 rpm) and incubated overnight at 37°C. DNA extraction was performed twice with an equal volume of phenol and twice with an equal volume of chloroform : isoamyl alcohol (24:1). For DNA precipitation, 1/10

volume of 3 M sodium acetate (pH 4.6) and 2 volumes of cold ethanol were added to the tube and inverted several times until the DNA precipitated. DNA was transferred to an eppendorf tube and washed twice with 70% ethanol. After desiccation, the DNA was dissolved in 0.1 ml of TE (10 mM Tris-HCl/0.1 mM EDTA). Gloves were used throughout the procedure, and phenol and chloroform were handled in a fume hood.

2.2.4. Electroelution of DNA (modification of method of Maniatis et al. 1982)

DNA was recovered from low melting point (LMP) agarose gel or agarose gel slices (Pharmacia) by electroelution after ethidium bromide (EtBr) staining and visualization with UV illumination. The dialysis tubing (Promega) used for electroelution was prepared by boiling for 10 minutes in one litre of 2% sodium bicarbonate/1 mM EDTA. After rinsing thoroughly in distilled water, the dialysis tubing was boiled for 10 minutes in distilled water, allowed to cool and stored at 4°C. The tubing was washed inside and outside with distilled water before use.

A slice of LMP agarose gel containing the DNA was excised and placed into pretreated dialysis tubing containing 0.5 x TBE (1 x TBE: 89 mM Tris base, 89 mM boric acid and 10 mM EDTA). DNA was electrophoresed out of the gel slice at 100V for 2-3 hours in 0.5 x TBE. The current direction was reversed for 2 minutes and the buffer containing the DNA was recovered from the dialysis tubing. DNA was recovered by ethanol precipitation.

## 2.3. PURIFICATION OF DNA

### 2.3.1. Phenol / Chloroform Extraction of DNA (Modification of method of Wyman and White, 1980)

Solutions of DNA were extracted with phenol/chloroform to remove proteins and other contaminants. Equal volume of phenol (TE saturated, 10 mM Tris-HCl, 1 mM EDTA) was added to the DNA solution which was then vortexed vigorously for 1 minute. After vortexing, the mixture was centrifuged for 5 minutes at full speed in an Eppendorf centrifuge. The upper aqueous phase which contained the DNA was removed leaving a white interface of denatured protein and the lower organic phase. Phenol extraction was followed by a phenol/chloroform extraction. An equal volume of phenol and chloroform were added to the DNA solution, which was then vigorously mixed and centrifuged for 5 minutes. The aqueous phase was removed and added to an equal volume of chloroform : isoamyl alcohol (24:1). The vortexing, centrifugation and aqueous phase removal steps were repeated once again. Following the extraction procedure, the DNA was then ethanol precipitated.

### 2.3.2. Ethanol Precipitation of DNA

To the DNA sample was added 1/10 volume of 3 M sodium acetate at pH 5.2, and 2 volumes of cold ethanol and the solution mixed well. The mixture was placed at -20°C overnight. Precipitated DNA was pelleted at 17,000g for 15 minutes and

washed once in 70% ethanol. After drying in air or under vacuum, the DNA was redissolved in $H_2O$ or TE.

2.3.3. Purification of PCR Product Using Prep-A-Gene

The protocol was obtained from Bio-Rad and the reagents used in this protocol were provided in the Prep-A-Gene kit.

Kit Components: Prep-A-Gene DNA Purification Matrix

Prep-A-Gene Binding Buffer

Prep-A-Gene Wash Buffer

Prep-A-Gene Elution Buffer

The PCR reaction was removed from beneath the paraffin oil to a clean tube. Three volumes of binding buffer were added to the PCR reaction and mixed briefly. 5 µl Prep-A-Gene matrix (5 µl will bind up to 1 µg of DNA) was added to the solution and the tube flicked. The solution was incubated at room temperature for 5-10 minutes ( frequent agitation during the binding step will assist DNA to bind to the Prep-A-Gene matrix ).The solution was centrifuged for 30 seconds at full speed and the supernatant removed. The pellet containing the bound DNA was rinsed by resuspending it in an amount of binding buffer equivalent to 50 times the amount of added matrix. The solution was centrifuged as above, and the rinse step was

repeated twice more. The pellet was washed three times with a 50 times matrix volume of wash buffer. After the last wash, all traces of liquid in the tube were carefully removed by recentrifugation. The matrix was dried in a speed-vacuum concentrator (model RH 40-11, SAVANT Instruments INC.) for a few minutes to remove all traces of ethanol. To elute the bound DNA, the matrix pellet was suspended in 30 µl of elution buffer (at least one pellet volume of elution buffer ) and incubated at 37-50°C for 5 minutes. The eluted DNA was transferred to a clean tube after centrifugation at full speed for 30 seconds.

## 2.4. POLYMERASE CHAIN REACTION (PCR) AND POLYACRYLAMIDE GEL ELECTROPHORESIS

In this study, the PCR technique was used for characterization of STRs (chapter 3), for physical mapping of STRs using human-mouse somatic cell hybrids and for genotyping STRs using CEPH reference families (chapter 4) and familial Mediterranean fever families (chapter 6).

### 2.4.1. PCR Reactions

2 x PCR Reaction Mix ( modified from Kogan et al., 1987)

| | |
|---|---|
| 33mM | $(NH_4)_2SO_4$ |
| 133mM | Tris-HCl (pH 8.8) |
| 20mM | β-mercaptoethanol (freshly added to incubation mix) |
| 13µM | EDTA |

| | |
|---|---|
| 0.34mg/ml | Bovine Serum Albumin (BSA) |
| 20%(v/v) | Dimethyl Sulphoxide (DMSO) |
| 3mM | dATP |
| 3mM | dGTP |
| 3mM | dCTP |
| 3mM | dTTP |

Incubation Reaction :

| | | |
|---|---|---|
| 5 | µl | 2 x PCR Reaction Mix |
| 1 | µl | $MgCl_2$ ( concentration range: 3 - 6.0 mM, need to optimize ) |
| 1 | µl | Oligo Primer Pair ( 150 ng/µl each oligo ) |
| 1 | µl | Template DNA (100 ng ) |
| 1 | µl | 1/10 diluted *Taq* Polymerase (0.5 units) with 0.25 µCi [$\alpha$-$^{32}$P] dCTP |

(If doing cold PCR, $^{32}$P-dCTP was not added.)

2.4.2. PCR Amplification Conditions

All PCR reactions were performed in a Perkin-Elmer Cetus Thermal Cycler. The PCR reaction components were as described above. The simple tandem repeats could be amplified by oligo primers recognising single copy DNA sequences flanking the repeat blocks. The PCR conditions for amplification of each STR marker were optimized using different concentrations of $MgCl_2$ and different cycling files. Most of the STR markers were amplified using a $MgCl_2$ concentration range from 3.5-5.0 mM. When the PCR incubation reaction was set up, one drop

of paraffin oil was added to the surface of the reaction. The PCR amplification was carried out with denaturation, annealing and extension for each cycle. The temperatures and times for annealing and extension were dependant on the primer pairs being used and the length of PCR product expected. The number of cycles was determined by the expected yield of PCR product. Most of the PCR amplifications were performed under the following conditions (cycling file 21): 94°C for 1 minute, 60°C for 1.5 minutes and 72°C for 1.5 minutes for a total of 10 cycles, followed by 94°C for 1 minute, 55°C for 1.5 minutes and 72°C for 1.5 minutes for a total of 25 cycles and a final extension at 72°C for 10 minutes.

For a $^{32}$P-incorporated PCR reaction, 3 volumes of formamide stop solution (95% deionized formamide, 1mM EDTA, 0.1% xylene cyanol FF, 0.1% Bromophenol blue) were added to the reaction, which was then denatured at 94°C for 3 minutes before loading on the polyacrylamide gel.

2.4.3. Polyacrylamide Gel Electrophoresis

Polyacrylamide gel mix comprised of 3.5-5% polyacrylamide (acrylamide: bisacrylamide = 19 : 1) in 1 x TBE with 7M urea. The mixture was filtered through Millipore filter paper and added to 0.1% fresh ammonium persulphate and 0.05% TEMED (N,N,N$^1$,N$^1$ -tetra-methylethylenediamine) before pouring the gel. The gel was poured and formed between glass plates separated by 0.25mm (for sequencing) or 0.4mm (for microsatellite genotyping). The gel was pre-electrophoresed in 1 x

TBE until the gel temperature reached 45-50°C. 3-6 µl of denatured sample was loaded per slot and the gel run at 2,000-2,500 volts based on the gel temperature (the best temperature is 50°C) for the appropriate time.

After electrophoresis, the gel was transferred to a filter paper, covered with plastic film and dried at 80°C for 2 hours using a 583 Gel Drier (Bio-Rad). The dried gel was exposed to X-0mat XK-1 film (Kodak).

# CHAPTER 3

## ISOLATION AND CHARACTERIZATION OF STR MARKERS

## 3.1. INTRODUCTION

Simple tandem repeats exhibit length polymorphisms. They are generally highly polymorphic and represent a rich source of informative genetic markers. For construction of a high-resolution genetic linkage map of human chromosome 16, two approaches were used to isolate STR markers. In the first stage of this study, the random isolation approach was exploited to generate STR markers from the chromosome 16 cosmid library (Stallings et al., 1990). The cosmid library was constructed from the human-mouse somatic cell hybrid CY18 (Callen, 1986), which contains chromosome 16 as the only human chromosome. After (AC)n repeats isolated from the random approach were mapped to 9 breakpoint intervals of chromosome 16, the second approach was used to generate repeat markers with which to fill in the deficient intervals or to isolate additional markers in regions of particular interest [(fragile sites *FRA16A* and *FRA16B*, disease gene regions of adult autosomal polycystic kidney disease (PKD1) (Reeders et al., 1988), familial mediterranean fever (FMF)(Aksentijevich et al., 1993b) and Batten disease (CLN3)(Mitchison et al., 1993)]. The second approach involved the isolation of STR markers from cosmid clones which were previously mapped to the above mentioned specific regions of chromosome 16 using the hybrid panel (Callen et al., 1992; Stallings et al., 1992).

In order to select the highly polymorphic STR markers for construction of genetic map of chromosome 16, the isolated STR repeats were characterized by determining allele size, allele frequency and heterozygosity.

## 3.2. MATERIALS AND METHODS

### 3.2.1. Isolation of STR Markers

#### 3.2.1.1. Subcloning

##### 3.2.1.1.1. Preparation of Human DNA Inserts and Vector DNA (for random isolation) (modification of method of Maniatis et al. 1989)

500 ng M13 mp 19 vector (New England Biolabs) was digested with *Sal* I in a total volume of 20 µl at the required temperature for 1 hour. The digestion was tested by running 1 µl of digested and undigested vector DNA sample side by side on a minigel which was stained with ethidium bromide (EtBr) and visualised under UV light. The concentration of M13 DNA was adjusted to 100 ng/ml with water. 10 µg DNA extracted from the chromosome 16 cosmid library was digested with *Sau*3AI (New England Biolabs). The digested DNA sample was checked on a minigel for complete digestion. The 200-500bp fragments selected from the cosmid DNA digestion were extracted by electroelution, phenol/chloroform extraction and ethanol precipitation. The final concentration of insert DNA was adjusted to 50 ng/ml with TE or water.

3.2.1.1.2 Partial Filling *Sau*3AI Site of Insert and *Sal* I Site of Vector

(modified method of Maniatis et al. 1989)

Solution: 10 x Fill-in Buffer

0.5 M Tris-HCl, pH 7.2

0.1 M $MgSO_4$

1 mM Dithiothreitol (DTT)

500 µg/ml Acetylated BSA

10 mM dCTP / 10mM dTTP

or

10 mM dGTP / 10mM dATP

In order to avoid polyinsert subcloning into the vector, additional dNTPs were incorporated into both the 3'-ends of the insert and the linearized vector (called partial fill-in, Figure 3-1). 20 µl of 100ng/µl *Sal* I cut M13 was added to 2 µl of 10 x partial fill-in buffer (dCTP/dTTP) and 1µl of Klenow enzyme (Amersham). 35 µl of chromosome 16 cosmid library DNA (200-500bp *Sau*3AI fragments, 30 ng/µl) was added to 2 µl of 10 x partial fill-in buffer (dGTP/dATP) and 1 µl of Klenow enzyme. Both solutions were incubated at 37°C for 30 minutes. Phenol/chloroform extraction and ethanol precipitation were then used to purify the DNA. After partial fill-in, C and T bases were added to the 3' end of the *Sal* I cut M13, and G and A bases were added to the 3' end of the *Sau*3AI cut insert. The remaining two base

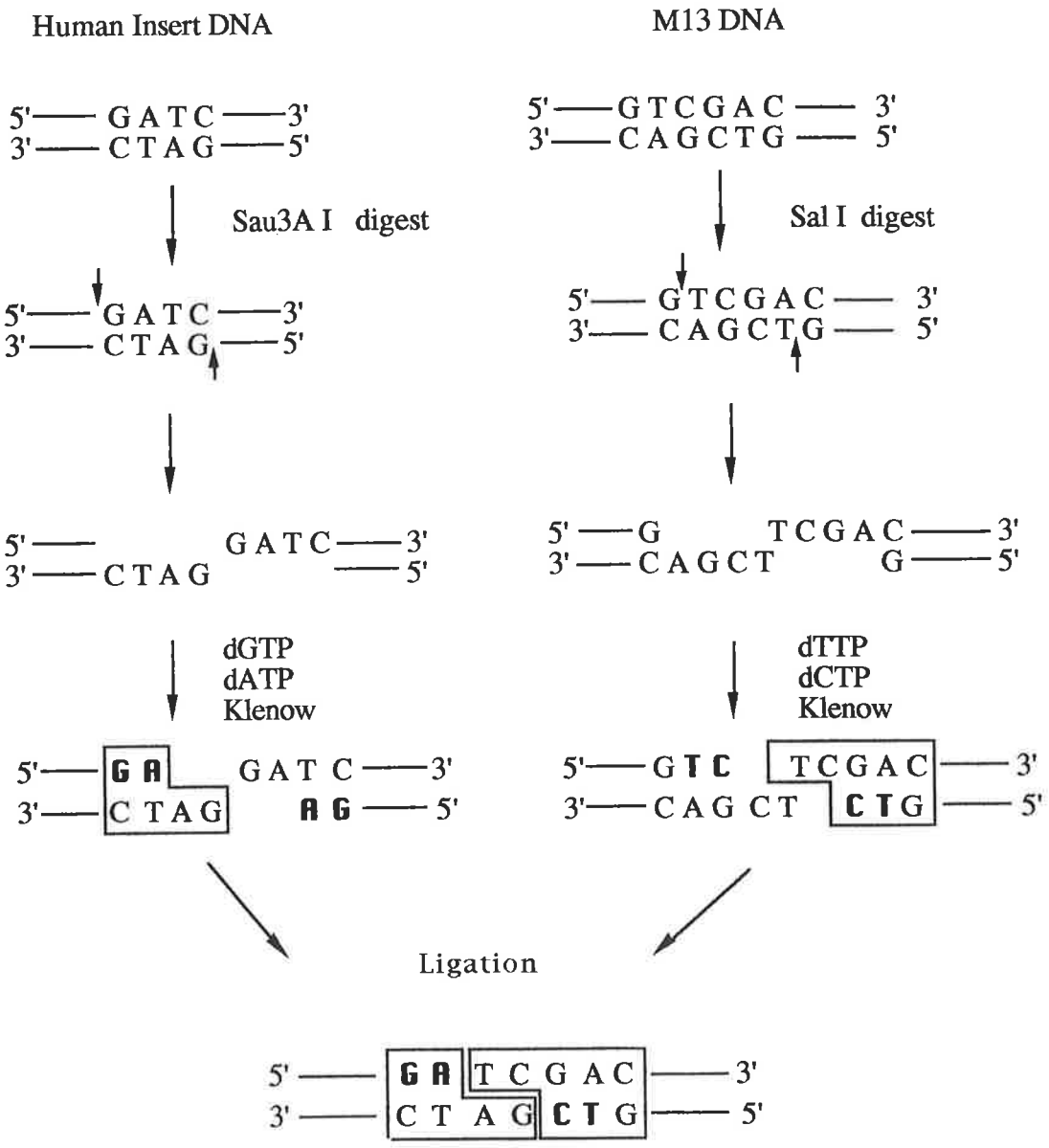



**Figure 3-1.** Partial fill-in the ends of human insert DNA and M13 vector DNA

pair 5' overhang allowed complimentary ligation of vector and insert DNAs and prevented their self-ligation.

3.2.1.1.3. $^{32}$P-Labelling of Dinucleotide Probes (product handbook of Amersham, UK)

100 ng poly(GT)n/(AC)n (Pharmacia) or 100 ng $(AG)_{30}/(TC)_{30}$ (synthesized by K. Holman using an Applied Biosystems 391 DNA Synthesizer) were denatured at 100°C for 2 minutes. The DNA was then labelled using 5 µl of 10 x buffer and 1 µl of Klenow enzyme from multiprime DNA labelling systems kit (RPN.1601Z Amersham), and 5 µl of [$\alpha$-$^{32}$P] dCTP (Amersham), in a total volume of 50 µl. The labelling reaction was incubated at 37°C for 30 minutes or at room temperature for 3 hours. The labelled DNA was denatured at 100°C for 10 minutes and cooled on ice for 5 minutes. The reaction was then ready for hybridization.

3.2.1.1.4. 5'End-$^{32}$P Labelling of Dinucleotide and Trinucleotide Probe (modified method of product handbook of Promega, Madison, USA)

2 µl of 50 ng/µl single stranded trinucleotide or 2 µl of 50 ng/µl $(AG)_{30}/(TC)_{30}$ (synthesized as mentioned above) was added to 1 µl of 10 x buffer (0.5 M Tris-HCl pH 9.5, 0.1 M $MgCl_2$, 50 mM DTT, 50% glycerol), 1 µl of [$\gamma$-$^{32}$P] dCTP, 1 µl of T4 DNA polynucleotide kinase and 5 µl of $H_2O$. The labelling reaction was

incubated at 37°C for 30 minutes. The reaction was then ready for hybridization. (This method was modified from the Promega method).

3.2.1.1.5. Hybridization (modification of method of Maniatis et al. 1989)

a) Hybridization with $^{32}$P-labelled Probes

The filter was prehybridized with a solution containing 0.5M $Na_2HPO_4/NaH_2PO_4$ (pH7.0) and 7% SDS in a 65°C water bath for 30 minutes to 1 hour. The $^{32}$P-labelled dinucleotide probe was added to the prehybridized filter. The hybridization was carried out in a 65°C water bath with shaking overnight. The filter was washed twice with 0.1% SDS and 2 x SSC in a 65°C for 15 minutes. The filter was exposed to X-0mat XK-1 film (Kodak) at -70°C in presence of two intensifying screens.

b) Hybridization with 5'end-$^{32}$P Labelled Probes

The filter was prehybridized with 20 ml of hybridization solution [ 5 x SSPE (sodium chloride, sodium dihydrogen orthophosphate and EDTA), 1% SDS, 200 ml of deionized formamide, 40g of dextran sulphate, 60 ml of $H_2O$ in a total volume of 400ml] at 30°C for 30 minutes. The 5' end-$^{32}$P labelled reaction was added to the filter and hybridized in a 30°C water bath with shaking overnight. The filter was washed twice with 5 x SSC and 1% SDS at room temperature for 2-4

minutes. The filter was then exposed to X-0mat XK-1 film (Kodak) at -70°C in presence of two intensifying screens.

3.2.1.1.6. Preparation of Human DNA Inserts and Vector DNA

(for specific isolation) (modification of method of Maniatis et al. 1989)

1 µg of M13 mp19 DNA (Amersham, UK), was digested with *Bam*HI and *Hinc*II, respectively, which cleave the polylinker, in a total volume of 50 µl at 37°C for 2 hours. The digestion was tested by running 1 µl of digested and undigested vector DNA samples side by side on a 0.8% agarose minigel which was stained with EtBr and visualised under UV light. The human DNA from the cosmids which were selected from specific regions of chromosome 16 was dot-blotted onto a nylon membrane filter with 1µl DNA for each cosmid clone. The filter was hybridized with $^{32}$P-labelled poly(GT)n/(AC)n, $^{32}$P-labelled $(AG)_{30}/(TC)_{30}$ or end-labelled trinucleotide probes at 65°C overnight. The filter was washed twice with 0.1% SDS and 2 x SSC at 65°C for 15 minutes. The filter was then exposed to X-0mat XK-1 film (Kodak) at -70°C. The positive cosmid clones were initially digested with several restriction enzymes. Electrophoresis was performed for each digestion and DNA was transferred from agarose gel to nitrocellulose filter. The filter was hybridized with $^{32}$P-labelled poly(GT)n/(AC)n (Pharmacia), $^{32}$P-labelled $(AG)_{30}/(TC)_{30}$ or $^{32}$P-end labelled trinucleotide probes. The enzyme which digested the repeat containing DNA fragment to 200-500bp was chosen for digestion of

cosmid DNA. The cosmid DNA was digested with the selected enzyme in a total volume of 20 µl at 37°C for 2 hours. The digestion was followed by heating at 65°C for 20 minutes to inactivate the enzyme. The reaction was directly used for ligation into M13.

3.2.1.1.7. Dephosphorylation of Vector DNA (product handbook of Epicentre Technologies, Madison, USA)

To prevent self-ligation of vector digested with a single restriction enzyme, the 5' terminal phosphate group was removed with HK$^{TM}$ phosphatase (Epicentre Technologies). When vector DNA was digested to completion in a total volume of 50 µl with an appropriate restriction enzyme, 6 µl of 1 unit/ul phosphatase, 7 µl of 10 x buffer and 7 µl of 50mM $CaCl_2$ were added into the digestion. The reaction was carried out at 30°C for 1 hour and then followed by heating at 65°C for 20 minutes to inactivate the phosphatase. The dephosphorylated M13 DNA was directly used for ligation.

3.2.1.1.8. Ligation Reaction (modification of the method from product handbook of Promega, Madison, USA)

Ligation reactions were carried out with a vector:insert molar ratio of approximately 1:3 or 1:10 to maximise intermolecular ligation rather than intramolecular ligation. For 20ng of linearized phosphatased M13 vector, 2 µl of 10

x ligation buffer, 1-2 units of T4 DNA ligase and insert DNA ( 5-50 ng ) were added and the reaction ( in a total volume of 20 µl) was incubated at 16°C overnight.

3.2.1.1.9. Transformation (modification of method of Maniatis et al. 1989)

*E. coli* strain BB4 cells were made competent with a method modified from Bullock et al. (1987). Stationary phase BB4 cells from an overnight culture were diluted 1:100 (v/v) into 20 ml 2 x TY medium ( 100 ml containing 16g bacto tryptone, 10g bacto yeast extract and 5g NaCl ). The cells were grown at 37°C with constant shaking for 2-2.5 hours. The cells were pelleted by centrifugation in a Jouan centrifuge at 3,000 rpm for 10 minutes and the cell pellet was then resuspended in 2 ml of ice cold fresh solution containing 0.1 M $CaCl_2$ and 20mM $MgCl_2$. The cells were ready for use after leaving on ice for 1 hour.

For the transformation reaction, 1 - 6 µl of ligation reaction was added to 200 µl of competent BB4 cells and left on ice for 40 minutes. The reaction was heated at 42°C for 2 minutes and placed back on ice for 5 minutes. 800 µl of SOC ( 2% bacto tryptone, 0.5% yeast extract, 10mM NaCl and 2.5 mM KCl) with 20mM $MgCl_2/SO_4$ and 20mM glucose were added into the competent cells. Before plating, the cells were transferred to the prewarmed solution with 8 ml of H-top agar (1,000 ml containing 10g bacto tryptone, 8g NaCl and 8g agar), 80 µl of 0.1 M isopropylthio-β-D-galactoside (IPTG), 80 µl of 6% 5-bromo-4-chloro-3-insolyl-β-

D-galactoside (X-gal) and 400 µl of fresh overnight cultured BB4 cells. The mixture was plated onto H-agar plates containing 10 µg / ml tetracycline. The plates were inverted and incubated at 37°C overnight (about 16 hours). The recombinant M13s were detected as colourless plaques.

3.2.1.2. Identification of Positive Clones (modification of method of Maniatis et al. 1989)

3.2.1.2.1. First Screen of the Positive Plaques

To lift the plaques, a nylon membrane filter was placed onto the plate surface (plates containing top agar were previously cooled to 4°C). The filter was left on the plate for 1 minute and transferred to 0.5 M NaOH solution with the plaque side up for 2 minutes to lyse host cells and denature the DNA. The filter was then transferred to 1 M Tris-HCl pH 7.5 for 2 minutes for neutralisation. The filter was washed in 2 x SSC and baked in a microwave oven at high heat for 50 seconds.

The filter was hybridized with $^{32}$P-labelled poly(GT)n/(AC)n, $^{32}$P-labelled $(AG)_{30}/(TC)_{30}$ or $^{32}$P-end labelled trinucleotide probes in a 65°C water bath with shaking overnight and washed twice with wash solution (0.1% SDS and 2 x SSC ) at 65°C for 15 minutes. The filter was then exposed to X-0mat XK-1 film (Kodak) at -70°C in the presence of two intensifying screens.

3.2.1.2.2. Second Screen of the Positive Plaques

To make the second screen plate, 9 ml H-top agar containing 400 µl fresh overnight cultured BB4 cells and 4 µl of 25 ug/ul tetracycline was poured onto an H plate. After the top agar set, the positive plaques identified by the first screen were picked from transformation plates and transferred onto a second screen plate using autoclaved barbecue bamboo sticks. The plates were inverted and incubated at 37°C overnight.

The plaques were lifted as mentioned above. The filters were hybridized with $^{32}$P-labelled poly(GT)/(AC)n, $^{32}$P-labelled (AG)$_{30}$/(TC)$_{30}$ or $^{32}$P-end labelled trinucleotide probes at 65°C overnight in a water bath with shaking. The filters were then washed twice in wash solution at 65°C for 15 minutes. The filters were exposed to X-0mat XK-1 film (Kodak) at -70°C in the presence of two intensifying screens.

3.2.1.2.3. Preparation of Single Stranded M13 from Positive Plaques (modified method of product handbook of Bio-Rad, Richmond, USA)

The positive M13 plaques identified from the second screen were picked with sterile barbecue bamboo sticks by touching the surface of the plaque and incubating in 1.5 ml of 2 x TY broth in a 10 ml tube. This was shaken at 37°C for 7-8 hours. The cells were transferred to an eppendorf tube and centrifuged for 5 minutes at

17,000g. The supernatant was carefully removed using a Pasteur pipette until the bacterial pellet started to move from the side of the tube. Centrifugation was repeated and the remaining supernatant carefully removed. 300 µl of 2.5 M NaCl with 20% polyethylene glycol (PEG) 6,000 was added to the supernatant and left at 4°C overnight. The precipitated phage were pelleted in an Eppendorf centrifuge for 15 minutes. The supernatant was removed and discarded. The residual PEG was carefully wiped away from the inner wall of the centrifuge tube using a tissue. The phage pellet was dissolved in 180 µl $H_2O$ and 20 µl of 10 x lysis buffer (100mM Tris-HCl pH 7.5, 1 mM EDTA, 5% SDS) was added. The suspension was incubated at 80°C for 10 minutes then allowed to cool to room temperature. Lysed phage were centrifuged at room temperature for 15 minutes at 17,000g. The phage DNA pellet was washed once in 70% ethanol and air-dried. The M13 DNA was redissolved in 20 µl $H_2O$ and stored at 4°C.

### 3.2.1.3. DNA Sequencing

#### 3.2.1.3.1. DNA Sequencing by Dideoxy Chain Termination

DNA sequencing used the dideoxy chain termination method (Sanger et al., 1977). This method can be used for sequencing single stranded DNA (M13) and double stranded DNA (e.g. plasmid, PCR product). In this study, manual sequencing was initially used, automated sequencing were used to sequence M13 clones and PCR products after the Department had the automated sequencer.

3.2.1.3.2. Manual Sequencing

Manual sequencing was carried out using the protocol of United States Biochemical Corporation (USB, Cleveland, USA). The sequencing reagents used were USB Sequencing Kit (No. 70700) and Perkin Elmer Cetus (PEC, Norwalk, USA) Sequencing Kit (No. N808-0001).

For each template, a single annealing reaction was in a total volume of 13 µl containing 5-7 µl (~1µg) of M13 DNA, 1 µl (0.5 pmol) of primer, 2 µl of annealing buffer and 3-5 µl $H_2O$. The reaction was heated at 70°C for 2 minutes and cooled to room temperature. While cooling, 4 µl of each termination reaction was aliquoted into an eppendorf tube and kept at room temperature. To the annealing reaction was added 2 ul of labelling mix, 0.5 ul of $\alpha$-$^{32}$P dCTP and 2 µl (3 units) of sequenase (1 in 8 dilution, USB) or Taq polymerase (2.5 units, Perkin-Elmer Cetus). This labelling reaction was mixed and incubated at 45°C for 5 minutes. 4 µl of labelling reaction was transferred to each termination tube and mixed. The termination reactions were incubated at 70°C for 5 minutes and then cooled to room temperature. To stop the reaction, 4 µl of formamide stop solution was added to the termination reaction. The termination reaction was heated at 95°C for 3 minutes before loading.

3 µl of the denatured sample was loaded on each slot of a 6% polyacrylamide sequencing gel ( bisacrylamide:acrylamide = 1:19 ). The gel was run at 2,000-2,500

volts for 4-5 hours until the xylene cyanol dye from the stop solution reached the bottom of the gel. The samples were denatured again and loaded a second time on the gel. When the bromophenol blue dye from the stop solution reached the bottom of the gel, the run was stopped.

The gel was lifted from the plate to 3 mm Whatman filter paper and covered with plastic wrap. The gel was dried using a Gel Drier (Bio-Rad). The dried gel was then exposed to X-Omat XK-1 film (Kodak) at room temperature overnight or at -70°C for a short time.

3.2.1.3.3. Automated Sequencing

The automated sequencing was performed by using an Applied Biosystems 373A DNA Automated Sequencer and *Taq* DyeDeoxy™ Terminator Cycle Sequencing Kit (No. 401113) and protocol. The automated sequencing was used to sequence single stranded M13 DNA and double stranded PCR product DNA. In order to get both flanking sequences for designing the primers, some single stranded M13s which were initially sequenced were amplified to double stranded by PCR.

a) Single Stranded M13 DNA Sequencing

5-9.5 µl (~1 µg) of M13 DNA was added to 1 µl (10ng) of primer, 9.5 µl of reaction mix (from the sequencing Kit) and $H_2O$ to a total volume of 20 µl. The

reaction was covered by one drop of paraffin oil and put in a Perkin-Elmer Cetus Thermal Cycler. The amplification was carried out at 96°C for 30 seconds, 50°C for 15 seconds, 60°C for 4 minutes for a total of 25 cycles.

The amplified DNA was purified using phenol / chloroform extraction recommended by Applied Biosystems (Foster City, USA). To the reaction was added 80 µl of $H_2O$ and 100 µl of chloroform to remove the paraffin oil. 100 µl of phenol : $H_2O$ : chloroform (68:18:14) was added to the reaction and mixed well. The mix was centrifuged at 14,000 rpm for 2 minutes and the lower organic phase discarded. The aqueous layer was re-extracted with a second 100 µl of phenol:$H_2O$:chloroform and centrifuged . The upper aqueous layer was transferred to a clean tube. 15 µl of 2M sodium acetate (pH 4.5) and 300 µl of absolute ethanol were added. The mixture was centrifuged at 14,000 rpm for 15 minutes at room temperature and washed with 70% ethanol. The DNA pellet was dried using a speed vacuum concentrator (model RH 40-11, SAVANT Instruments INC). The DNA was then ready for sequencing.

b) PCR Product Sequencing

A PCR product of single stranded M13 DNA was obtained by amplifying at 94°C for 1 minute, 60°C for 1.5 minutes and 72°C for 1.5 minutes for a total of 10 cycles, followed by 94°C for 1 minute, 55°C for 1.5 minutes and 72°C for 1.5 minutes for a total of 25 cycles, and extended at 72°C for 10 minutes.

The PCR product was purified using Prep-A-Gene (BioRad) as described above. To the double stranded PCR product DNA was added 1 µl of 20 ng/µl primer (plasmid $p^{UC19}$ forward or reverse primer, or specific primer synthesised from previous M13 sequencing), 9.5 µl of reaction mix and $H_2O$ to 20 µl. The reaction cycling was the same as for the M13 sequencing.

After purification by phenol/chloroform extraction, the DNA was ready for sequencing.

c) Loading and Running the Sequencing Gel

To each sample was added 4 µl of formamide solution (deionized formamide : 50 mM EDTA pH 8.0 = 5 : 1) and mixed well. The solution was centrifuged briefly to collect all the liquid at the bottom of the tube. Before loading, the sample was heated at 90°C for 2 minutes. 3 µl of each sample was loaded on the 6% gel (acrylamide : bisacrylamide).

3.2.1.4. Designing of Oligo Primers

3.2.1.4.1. Criteria for Designing Oligo Primers

Each oligo primer should be about 25bp (20-30bp) long. The GC content in primers was 50% and there were usually no stretches of more than four consecutive purines. In addition, the primer pairs were checked at their 3' ends to avoid the

possibility of primer-dimer formation. The size of the PCR product was between 100-250bp.

3.2.1.4.2. Synthesizing and Processing Oligo Primers

Oligodeoxynucleotide primers were synthesized by K. Holman using an Applied Biosystems 391 DNA Synthesizer. After synthesizing, the oligo primers were deprotected and cleaved from the column using concentrated ammonium hydroxide, and left in a 55°C water bath overnight. For desalting or purifying the oligo primers, a 10 x volume of N-butanol was added to the oligo solution which was then, vortexed and centrifuged at 1,600g for 30 minutes. The supernatant was discarded. The N-butanol purification was repeated once more. The oligo primer DNA pellet was dried in a speed vacuum concentrator and diluted in 400 µl $H_2O$. The oligo primer DNA was quantitated and diluted to 1mg/ml, and stored at -20°C.

3.2.2. Characterization of STR Markers

3.2.2.1. Determination of Allele Frequency and Heterozygosity by PCR

Allele frequencies and heterozygosities of STR markers were determined by PCR of DNAs from 80-100 unrelated Caucasian individuals and/or CEPH (Centre d'Etude du Polymorphisme Humain, Paris; Dausset et al., 1990) parent DNAs. PCR products were loaded on 5% polyacrylamide gels. The alleles differing by 2bp were distinguished after electrophoresis and exposure to X-ray film. The expected

heterozygosities of the STR markers were obtained by the computer program PIC [This program can calculate the expected heterozygosity and PIC (polymorphic information content) using allele frequencies.].

## 3.3. RESULTS

### 3.3.1. Random Isolation

From the random isolation of the human chromosome 16 cosmid library, sixty-three (AC)n/(GT)n positive M13 clones were sequenced. Oligo primers were designed to amplify 25 (AC)n repeats. Twelve of these (48% ) (Table 3-1 and Table 3-2) were characterised (Table 3-7) and physically mapped to 9 breakpoint intervals of chromosome 16 (Figure 4-3). Three (25%) of these 12 (AC)n repeats were independently identical to other three (AC)n repeats: 16AC7.46 was identical to 16AC7.11, 16AC7.42 and 16AC7.55 were identical to 16AC1.18 and 16AC1.1, respectively. 16AC1.18 and 16AC1.1 were previously isolated from the chromosome 16 cosmid library (Thompson et al., 1992). Ten (40% ) of the 25 (AC)n repeats were mouse repeats(Table 3-3). The human chromosome 16 cosmid library (Stallings, et al., 1990) was constructed from the human-mouse somatic cell hybrid CY18 (Callen, 1986). This library contains about 7% mouse genomic DNA (Stallings, et al., 1990), therefore, the mouse repeats were isolated at the same time the human repeats were isolated. Three repeats (12% ) failed to amplify by PCR.

**Table 3-1.** (AC)n Repeats Isolated from the Chromosome 16 Cosmid Library

| Locus | Clone name | Repeat sequence | PCR product size in bp |
|---|---|---|---|
| D16S288 | AC7.1 | AGG(GT)$_{22}$ACA | 160 |
| D16S317 | AC7.9 | GAC(GT)$_4$CCT(GT)$_4$TAT(GT)$_{13}$CAG | 136 |
| D16S319 | AC7.14 | (T)$_6$GGGTTT(GT)$_{17}$GG(T)$_9$G(T)$_8$ | 155 |
| | AC7.15<br>AC6.7* | GGC(GT)$_{21}$ATT | 126 |
| | AC7.22 | AAT(GT)$_6$AT(GT)$_3$N$_{16}$(GTTT)$_6$GTT | 168 |
| D16S308 | AC7.42<br>AC1.18 | TTT(GT)$_{27}$GAG | 163 |
| D16S289 | AC7.46<br>AC7.11 | ATG(GT)$_{22}$ACA | 162 |
| D16S300 | AC7.55<br>AC1.1 | GTT(GA)$_{12}$(GT)$_{24}$(GA)$_5$GCA<br>(GT)$_6$ATGGCA(GT)$_6$ATA | 171 |
| | AC7.59 | TAT(GT)$_{10}$GC(GT)$_7$TTA | 173 |
| D16S318 | AC8.20 | GAG(GT)$_{22}$ATG | 144 |
| D16S531 | AC8.15 | ACC(GT)$_{31}$AGG | 137 |
| D16S320 | AC8.52 | TGC(GT)$_{24}$GAG | 169 |

* The primer pair designed from this clone failed to give satisfactory results with either mouse or human genomic DNA

**Table 3-2.** Primer Sequences of 12 STR Markers Isolated from the Random Isolation

| Repeat | Forward Primer (5'- 3') | Repeat Sequence | Reverse Primer (5'- 3') |
|---|---|---|---|
| AC7.1 | TAAATCACAGACTCCCTTGGCT | AGG(GT)$_{22}$ACA | ATGAGGAGCAGAGGTGTCTCTT |
| AC7.9 | AGGAGTACACAGATGCACCAGACCT | GAC(GT)$_4$CCT(GT)$_4$ TAT(GT)$_{13}$CAG | CCTTTGATACCCTGGCTTTGCAGTC |
| AC7.14 | ATCGCGCCCAACCCACTATT | TTT(GT)$_{17}$GG(T)$_8$ | GTGACAAGAGTGAAACTCCG |
| AC7.22 | ACCATGCCCAGCGTAGGTAATT | AAT(GT)$_6$AT(GT)$_3$ N$_{16}$(GTTT)$_6$TTG | ATCTACAGAAATGTTTGAGACC |
| AC7.42 | CAGCCAGGGTAGTAAGGCTAGACCT | TTT(GT)$_{27}$GAG | TGGGTGGCAGAGTGAGACCCTGCT |
| AC7.46 AC7.11 | CACCACTTATCATTTCTTCCAAGCTGTG | ATG(GT)$_{22}$ACA | AGTTGGAGGAAGAGAAGCAG |
| AC7.55 | AGCCAAGCTAGTTAATTTGTGGTCCCA | GTT(GA)$_{12}$(GT)$_{24}$ (GA)$_5$GCA(GT)$_6$ATG GCA(GT)$_6$ATA | AAGCGCCTGGGAGAGAGCAAGCTAT |
| AC7.59 | CTGAATATTTCTGTGCCTAGGTTTC | TAT(GT)$_{10}$GC(GT)$_7$TTA | CAACAGAGTGAGACCCTGTCTCTATA |
| AC8.15 | TAGAGTCGTTCTCACACACC | ACC(GT)$_{31}$AGG | TCATCTACAGGCATGGCACAAGC |
| AC8.20 | CTGTGGTGTCACATCAGGAA | GAG(GT)$_{22}$ATG | GACTACACATGAACCGATTG |
| AC8.52 | AGTCTGAGAGACATCCAGGT | TGC(GT)$_{24}$GAG | GTGATATCAGTCAGTCCTGT |

**Table 3-3.** The Mouse (AC)n Repeats

| Clone name | Repeat Sequence | PCR product size in bp |
|---|---|---|
| AC7.5 | GAG(GT)$_{16}$TTA | 112 |
| AC7.13 | GGG(GT)$_{18}$CT(GT)$_{5}$ATG | 241 |
| AC7.17 | AAT(GT)$_{22}$GCT | 175 |
| AC7.21 | ATT(GT)$_{15}$GCT | 125 |
| AC7.33 | TTG(CT)$_{16}$(GT)$_{24}$TTC | 148 |
| AC7.38 | TGA(GT)$_{13}$AT(GT)$_{13}$(GA)$_{17}$TGC | 155 |
| AC7.47 | CAT(GT)$_{26}$(GA)$_{18}$AAT | 225 |
| AC8.25 | TTT(GT)$_{2}$G(GT)$_{25}$ | 180 |
| AC8.31 | GAA(GT)$_{23}$ACA | 173 |
| AC9.2 | ACT(GT)$_{10}$CGTGCG(GT)$_{9}$AAG | 231 |

For 38 sequenced clones, primers were not designed due to 1) (AC)n were too short ( n < 16 ) or absent from available sequence; 2) (AC)n repeats were flanked by repetitive sequences; 3) clones were too close to the cloning site to design primers.

3.3.2. Specific Isolation

Seventy-four cosmid clones (shown in Table 3-4) of human chromosome 16 were selected from specific regions (deficient intervals, disease gene regions: adult autosomal polycystic kidney disease (PKD1) (Reeders et al., 1988; Somlo et al., 1993), familial mediterranean fever (FMF) (Aksentijevich et al., 1993b) and Batten disease (CLN3) (Mitchison et al., 1993; Yan et al., 1993) and fragile site regions: [*FRA16A* (Sutherland, 1979) and *FRA16B* (Sutherland, 1993)]. Fifty-eight cosmids which were (AC)n or (AG)n positive were subcloned into M13. For 30 (52% ) repeats, oligo primers were designed (Table 3-5 ). Three cosmids were also shown to be (CCG)n positive, of which 2 were subcloned into M13 and sequenced. (CCG)n was only found in one clone with 10 repeat units. In two instances, identical (AC)n repeats were isolated from cosmids which were subsequently shown to overlap: 1) an (AC)20 repeat was independently isolated from two cosmids 302H7 and PKDcos5 which were previously mapped in the PKD1 region. These two cosmids overlapped by ~8 kb. This repeat was also identical to a previously isolated repeat 16ACVK5 (*D16S94* ) which was isolated from a λ-clone, VK5, in the PKD1 region (Hyland et al., 1990). 2). a compound repeat

**Table 3-4.** Screening Simple Tandem Repeats from Cosmid Clones of Human Chromosome 16

| 58 Cosmids with positive (AC)n, (AG)n or (CCG)n | | | | |
|---|---|---|---|---|
| 308G7 | 302H7 | PKDcos5 | 77E8 | 310G9 |
| 325B10 | 2H2 | 13H1 | 301G12 | 311E7 |
| 28D3 | 49E7 | 52C5 | CRI.0114 | 45G5 |
| 1E5 | 25H11 | 33A4 | 40A7 | CRI.066 |
| 80B3 | 302E3 | 19F8 | 12E7 | 10F5 |
| 47G9 | 12F8 | 10B3 | 305B7 | 16F8 |
| 69F12 | 39C8 | 49C4 | 65C5 | 67G12 |
| 80H3 | 312C6 | 33G11 | 301B4 | 18A3 |
| 306E12 | 16F1 | CRI.015 | 305D6 | 305E9 |
| 311D8 | 323H4 | 51A4 | 36H4 | 51G1 |
| 325G11 | 26E3 | CRI.043 | 302G7 | 11C2 |
| 330F12 | 60B8 | 16F2 | | |
| 16 Negative Cosmids | | | | |
| 301E3 | 10B2 | 30C1 | 302B4 | 10B8 |
| 305F6 | 80A9 | 315F5 | 303F2 | 5A4 |
| CRI.095 | CRI.0101 | CJ52.27 | CJ52.94T1 | CJ52.1 |
| CJ52.10 | | | | |

**Table 3-5.** Primer Sequences of 32 STRs Isolated from the Specific Regions

| Repeat | Forward Primer (5'→3') | Repeat Sequence | Reverse Primer (5'→3') |
|---|---|---|---|
| AC308G7 | CCTCAGGCGCGAGACGGGCT | GCT$(GT)_{10}(GA)_{21}$ $(GT)_{21}$CGG | AGCTAGGAAGCGGGTCCTCGG |
| AC13H1 | CACAGTGTTCCATTTTCTC | CTC$(GT)_{19}$AGN | GTAGCAGAGANAGACTTTG |
| AC301G12 | TGGACTCATGGGTCAGGACTTAAT | CTG$(GT)_{11}$ACT $(GT)_{4}$GGT | TCGTGGTCTTGTTGTAATTGAGTC |
| ACCRI.0114 | AATTCAAGGGAGGCTCATGTG | TGG$(GT)_{6}$GGTGG $(GT)_{3}$GG$(GT)_{14}$TCA | CACTCCTCCCTCTATGTTATG |
| AC45G5 | GCTGCCATCCTAAGCTTTGGTTAGG | TAT$(GT)_{22}$ATG | AGGTCGACTCTAGAGGATCTTCAAC |
| AC1E5 | CTGTTCCCACCTTTCCCACAATCAGC | GGC$(GT)_{17}$TCT | TGTCTCCCTTNAGATGGCAGGGCTC |
| AC40A7 | GGAGAAGGCTAAAGATTCATTTGTGAG | GAA$(GT)_{19}$TGG | TGCAACAGGTAACTGTAACTCTATTCC |
| AC25H11 | AACTAGATATGCCCATACGATTGC | TCT$(GT)_{7}$GC $(GT)_{9}$TTA | CTAAGTAACTAATGCCTAGCTCCA |
| AC33A4 | GTTCTTGAACCTCCCCACTTTACCCTTAG | CAG$(GT)_{24}$AT $(GT)_{5}$AT$(GT)_{3}$ACA | TCAGTGNCCATGGTGTTGTTCCCCTG |
| ACCRI.066 | AGCTCAGGAGTTCAAGACCAGCCTGTGC | TCT$(AT)_{5}$TAA $(AC)_{22}$CAC | CAAGCTTGCATGCCTGCAGGTCCTAAC |

**Table 3-5.** Continued (1)

| Repeat | Forward Primer (5'→3') | Repeat Sequence | Reverse Primer (5'→3') |
|---|---|---|---|
| AC80B3 | CTCTCCCCATAGAAGGCATC | AGC$(GT)_{15}$AAG | GGATAGAGACGTTCTCTTAA |
| AC302E3 | ACAATCCAGCCTGGGTGACAGAGTG | CTC$(A)_{13}$ $(GT)_{14}$ATG | GTAATTGCCATATTTCATTTAGGAG |
| AC26E3B | AAATGCGTGTGTGCACCC | GCA$(GT)_{12}$ $(GC)_{4}$GTG | GCATGCATGCATGTGTATGC |
| AC10F5 | GACAGCCTCAAATGAAATATAACAC | CAT$(GT)_{14}(GC)_{3}$ $(GT)_{7}$TTC | CTCTCAGCTAGGGTAGTTGTTTATA |
| AC12F8 | CAGAAGCAGCAACATTACATGGTCCAGG | GGA$(GT)_{18}$ATC | CCACGCATAAGTGCTAGCTGAGCATATC |
| AC10B3 | TGGTCCTGATGTTTTAATTTAGCAATCCAG | CTC$(GT)_{16}$CTC | CATATTTCCATCAGGAATGTTAT |
| AC39C8 | ACACAAATTCCTAAAGCGTTTTTC | TTC$(GT)_{13}$GCG TGC$(GT)_{2}GC(GT)_{6}$AGA | GAGCCCAGGAGTTCGAGACCA |
| AC69F12 | ATATCCATCACCATTCATCCTTAC | TCT$(GT)_{21}$GGT | CTAAGGAAAGGTAGGACCTTGGT |
| AC8.21 | GAGAGTGAGAATTCCTAGCTTTGG | GAT$(GT)_{21}$ATG | CACGCTACAATCGCTACCACC |
| AC67G12 | AGCTATGTTATAAGCCAGAGA | GAA$(GT)_{5}$TT $(GTGTGC)_{4}(GT)_{21}$ AT$(GT)_{3}$GCA | AGACTCCTGCCTTCTGTGATAAACTC |

**Table 3-5.** Continued (2)

| Repeat | Forward Primer (5'→3') | Repeat Sequence | Reverse Primer (5'→3') |
|---|---|---|---|
| AC80H3 | ACATACAAAGTCTCTGGAGGAGTG | TGC(GT)$_{21}$AGA | CCACATCCTCATATCTACAACTCT |
| AC33G11 | GCATTCTAACGCAAAATTGAACTC | TTT(GT)$_{16}$TCA | TGCCTTGCACATGCCTGTTAATGG |
| ACCRI.015 | TTGCAGTCTTCATGCCTAAC | TAT(GT)$_{15}$ATA | GGTCCTCAAAGGGCCCTGG |
| AC305D6 | TAGGTCATGTTCATGCCACTACAC | TAT(GT)$_{13}$CAG | TGATTTAAAGGCACAGGAATATAT |
| AC305E9 | GATCTTCGGAATACTACTCGT | TGA(GT)$_{14}$ATG (TA)$_7$(T)$_{10}$GAG | CCATTGCACTCAGCCTGGGTGAC |
| AC51G1 | CACATGTTGATTCTCATTTACCTG | CGC(GT)$_{20}$ATA | TCTAAAGAAATCCTGTGAGAG |
| AC51A4 | GTGGATTTCAGACTTACCAACCTC | TCT(GT)$_{21}$CTG | ATTAGCAATGCTGGGTTTCCAGAG |
| AC323H4 | GAGGACATTTATTCATAGGAAGGAGG | GGT(GT)$_{25}$CGC | CGACTCTAGAGGATCCCAGACTTCG |
| AC325G11 | CAGGTAATGAAATACTCTAATTCAGTAAAT | GAT(GT)$_{11}$ AT(GT)$_{11}$(T)$_{15}$ | TGACTCTAGCCTGGCGACAGACTCAC |

16AC305E9 [(GT)14+(AT)7+(T)9 ] isolated from cosmid 305E9 was identical to the repeat isolated from cosmid 311D9. These two cosmids overlapped by ~15kb. Two repeats (16AC1E5 and 16AC39C8) detected two loci on chromosome 16, respectively. 2 repeats (16AC67G12 and 16ACCRI.066) failed to amplify by PCR due to the primer sequences being part of an *Alu* sequence. 16AC325G11 and 16AC302E3 failed to produce a satisfactory PCR product because the (GT)n repeat was adjacent to poly(T) or poly(A). Twenty-eight (48%) clones were discarded for the same kinds of reasons as previously described in the random isolation. Thirty repeats isolated from the second approach were shown in Table 3-6.

### 3.3.3. Characterization of STR Markers

#### 3.3.3.1. Allele Frequencies and Heterozygosities

Simple tandem repeats exhibit length polymorphism in the size of the amplified fragments. Alleles of (AC)n repeats always differ in size by multiples of 2 bases. Allele frequencies and heterozygosities were determined by PCR of DNAs from 80-100 unrelated Caucasian individuals and/or CEPH parents. The number of alleles found from 32 markers [ 30 markers isolated by the candidate and 2 markers (*D16S397* and *D16S398* ) provided by Dr. J. L. Weber (personal communication, 1991)] ranged from 2-17. For 4 markers [*D16S453* (AC301G12), *D16S332* (AC305D6), *D16S370* (25H11) and *D16S317* (AC7.9 )], 2 or 3 alleles were found. The rest of the markers all had in excess of 5 alleles. For most of these markers the major alleles were clustered in

**Table 3-6.** Simple Tandem Repeats Isolated from the Specific Regions

| Repeat | Cosmid | Cloned Fragment | Repeat Sequence | PCR Product Size of Cloned Fragment |
|---|---|---|---|---|
| 16AC308G7 | 308G7(D16S525) | Alu I | GCT$(GT)_{10}(GA)_{21}$ $(GT)_{21}$CGG | 157bp |
| | 302H7 | Hae III | CTC$(GT)_{20}$TAC | |
| | PKDcos5 | Hae III | | |
| 16ACVK5 | VK5(D16S94) | | | 86bp |
| 16AC13H1 | 13H1(D16S523) | Sau3A I | CTC$(GT)_{19}$AGN | 85bp |
| 16AC301G12 | 301G12(D16S453) | Sau3A I | CTG$(GT)_{11}$ACT $(GT)_4$GGT | 127bp |
| 16ACCRI.0114 | CRI-0114(D16S49) | HaeIII | TGG$(GT)_6$GGTGG $(GT)_3$GG$(GT)_{14}$TCA | 130bp |
| 16AC45G5 | 45G5(D16S454) | Sau3A I | TAT$(GT)_{22}$ATG | 183bp |
| 16AC1E5 | 1E5 | Sau3A I | GGC$(AC)_{17}$TCT | 133bp |
| 16AC40A7 | 40A7(D16S524) | Alu I | GAA$(GT)_{19}$TGG | 161bp |
| 16AC33A4 | 33A4(D16S370) | Sau3A I | CAG$(GT)_{24}$AT$(GT)_5$ AT$(GT)_3$ACA | 140bp |
| 16AC25H11 | 25H11(D16S370) | Sau3A I | TCT$(GT)_7$GC$(GT)_9$TTA | 96bp |
| 16ACCRI.066 | CRI-066(D16S42) | Alu I | TCT$(AT)_5$TAA $(AC)_{22}$CAC | 142bp |
| 16AC80B3 | 80B3(D16S383) | Rsa I | AGC$(GT)_{15}$A$(TG)_4$CAT | 152bp |
| 16AC302E3 | 302E3 | Sau3A I | CTC$(A)_{13}(GT)_{14}$ATG | 113bp |
| 16AC26E3B | 26E3B(D16S359) | Alu I | GCA$(GT)_{12}(GC)_4$GTG | 92bp |
| 16AC10F5 | 10F5(D16S390) | Hpa II | CAT$(GT)_{14}(GC)_3$ $(GT)_7$TTC | 185bp |
| 16AC69F12 | 69F12(D16S451) | Sau3A I | TCT$(GT)_{21}$GGT | 103bp |
| 16AC12F8 | 12F8(D16S347) | Alu I | GGA$(GT)_{18}$ATC | 186bp |
| 16AC39C8 | 39C8 | Sau3A I | TTC$(GT)_{13}$GCGTGC $(GT)_2$GC$(GT)_6$ | 124bp |

**Table 3-6.** Continued

| Repeat | Cosmid | Cloned Fragment | Repeat Sequence | PCR Product Size of Cloned Fragment |
|---|---|---|---|---|
| 16AC67G12 | 67G12(D16S153) | Alu I | GAA(GT)$_5$TT (GTGTGC)$_4$(GT)$_{21}$ AT(GT)$_3$GCA | 148bp |
| 16AC8.21 | (D16S346) | Alu I | GAT(GT)$_{24}$TTG | 109bp |
| 16AC80H3 | 80H3(D16S450) | Rsa I | TGC(GT)$_{21}$AGA | 101bp |
| 16AC33G11 | 33G11(D16S395) | Sau3A I | TTT(GT)$_{16}$TCA | 112bp |
| 16AC10B3 | 10B3(D16S389) | Sau3A I | CTC(GT)$_{16}$CTA | 99bp |
| 16ACCRI.015 | CRI.015(D16S40) | Alu I/Rsa I | TAT(GT)$_{15}$ATA | 118bp |
| 16AC305D6 | 305D6(D16S332) | Alu I | TAT(GT)$_{13}$CAG | 126bp |
| 16AC305E9 | 305E9(D16S392) 311D9 | Alu I | TGA(GT)$_{14}$ATGT(AT)$_7$ (T)$_9$GAG | 121bp |
| 16AC51G1 | 51G1(D16S363) | Sau3A I | CGC(GT)$_{20}$TTA | 254bp |
| 16AC323H4 | 323H4(D16S393) | Alu I | GGT(GT)$_{25}$CGC | 152bp |
| 16AC51A4 | 51A4(D16S449) | Alu I | TCT(GT)$_{21}$CTG | 193bp |
| 16AC325G11 | 325G11(D16S358) | Taq I | GAT(GT)$_{11}$AT (GT)$_{11}$(T)$_{15}$AGG | 161bp |

size, within about 10bp, on either side of the predominant allele. The expected heterozygosities of these 32 markers ranged from 0.11 to 0.91. Twenty-six (81%) of markers had heterozygosities greater than 0.50 [17 (65%) had a heterozygosity greater than 0.70, and 5 (19%) had a heterozygosity of 0.68 or 0.69]. The number of alleles and the expected heterozygosities of the 32 markers are shown in Table 3-7.

In addition, one marker (*D16S399*, mfd186), which was thought to localize to chromosome 16, was provided by Dr. J. L. Weber (personal communication, 1991) was further characterised by the candidate (Table 3-7). This marker did not physically localize to chromosome 16 and was found to be localized to human chromosome 1q after linkage analysis.

3.3.3.2. Informativeness of STR Markers

Thirty-two STR markers were divided into three categories (Table 3-8) according to the classification rules which Weber described previously (Weber, 1990a): 1) 22 perfect repeats ( 69% ). 2) 5 imperfect repeats (15%). 3) 5 compound repeats (15%). The 32 repeats were also divided into two categories according to the maximum number of perfect repeat units which each marker contained (Table 3-9). All 22 (69%) of the markers with perfect repeat units ≥ 16 had heterozygosities greater than 0.50 (19 of them had heterozygosity ≥ 0.68). Ten markers (31%) with perfect repeat units 9-15 had heterozygosities 0.11-0.80 (3 markers with perfect repeats units of 14 had heterozygosities 0.68-0.80). As Figure 3-2 showed, the heterozygosity generally

**Table 3-7.** Characterisation of 32 STR Markers of Chromosome 16

| Locus | Clone | Repeat | No. Alleles (Size Range) | Het | Genotype* 133101 | Genotype* 133102 | Location |
|---|---|---|---|---|---|---|---|
| D16S525 | 308G7 | (GT)10(GA)21(GT)21 | 17(143-175) | 0.91 | 159,161 | 149,165 | p13.3 |
| D16S523 | 13H1 | (GT)19+(GT)3 | 5(77-87) | 0.68 | 77, 77 | 77, 85 | p13.3 |
| D16S453 | 301G12 | (GT)11+(GT)4 | 2(125-127) | 0.45 | 127,127 | 125,127 | p13.3 |
| D16S49 | CRI.0114 | (GT)6+(GT)3+(GT)14 | 7(108-132) | 0.68 | 110,130 | 108,130 | p13.11-p13.2 |
| D16S454 | 45G5 | (GT)22 | 10(173-193) | 0.75 | 175,175 | 175,179 | p13.11-p13.2 |
| D16S524 | 40A7 | (GT)19 | 11(143-169) | 0.76 | 163,169 | 143,155 | p12.1-p13.11 |
| D16S452 | 33A4 | (GT)24 | 7(132-144) | 0.68 | 140,140 | 134,140 | p12.1-p13.11 |
| D16S370 | 25H11 | (GT)7+(GT)9 | 2(96-112) | 0.30 | 112,112 | 96, 96 | p12.1-p13.11 |
| D16S319 | 7.14 | (T)6+(GT)17+(T)9 | 6(145-159) | 0.52 | 149,157 | 149,149 | p12.1-p13.11 |
| D16S288 | 7.1 | (GT)22 | 6(154-166) | 0.73 | 160,160 | 158,160 | p11.2-p12.1 |
| D16S383 | 80B3 | (GT)15 | 5(148-156) | 0.45 | 148,152 | 152,152 | p11.2-p12.1 |
| D16S359 | 26E3B | (GT)12+(GC)4 | 5(88-104) | 0.42 | 92, 96 | 92, 92 | q21.1-q13 |

* genotype of individuals from CEPH reference pedigree # 1331

**Table 3-7.** Continued (1)

| Locus | Clone | Repeat | No. Alleles (Size Range) | Het | Genotype* 133101 | Genotype* 133102 | Location |
|---|---|---|---|---|---|---|---|
| D16S390 | 10F5 | (GT)14+(GC)3+(GT)7 | 11(177-197) | 0.80 | 181,185 | 177,179 | q12.1-q13 |
| D16S531 | 8.15 | (GT)31 | 17(111-149) | 0.86 | 125,149 | 119,123 | q12.1-q13 |
| D16S317 | 7.9 | (GT)4+(GT)4+(GT)13 | 3(134-138) | 0.16 | 136,136 | 136,138 | q12.1-q13 |
| D16S320 | 8.52 | (GT)24 | 12(151-175) | 0.86 | 163,173 | 163,169 | q12.1-q13 |
| D16S389 | 10B3 | (GT)16 | 11(88-114) | 0.77 | 98, 110 | 98, 100 | q21 |
| D16S451 | 69F12 | (GT)21 | 10(83-113) | 0.84 | 105,111 | 111,113 | q21 |
| D16S347 | 12F8 | (GT)18 | 9(186-204) | 0.76 | 186,190 | 186,190 | q22.1 |
| D16S398 | MFD168** | (GT)22 | 10(224-242) | 0.85 | 226,228 | 230,234 | q22.1 |
| D16S397 | MFD98** | (GT)22 | 9(179-195) | 0.83 | 185,189 | 187,189 | q22.1 |
| D16S318 | 8.20 | (GT)22 | 6(134-152) | 0.54 | 136,144 | 144,146 | q22.1 |
| D16S522 | 8.21 | (GT)21 | 8(103-119) | 0.69 | 105,109 | 105,113 | q22.1-q23.1 |
| D15S450 | 80H3 | (GT)21 | 8(83-103) | 0.52 | 85, 85 | 85, 93 | q22.1-q23.1 |

** Provided by Dr. J.L.Weber (personal communication, 1991), characterised by the candidate.

**Table 3-7.** Continued (2)

| Locus | Clone | Repeat | No. Alleles (Size Range) | Het | Genotype* 133101 | Genotype* 133102 | Location |
|---|---|---|---|---|---|---|---|
| D16S395 | 33G11 | (GT)16 | 9(98-124) | 0.69 | 98, 98 | 98, 114 | q23.1-q24.1 |
| D16S289 | 7.46 | (GT)21 | 8(156-172) | 0.77 | 164,164 | 162,164 | q23.1-q24.1 |
| D16S40 | CRI.015 | (GT)15 | 5(118-128) | 0.11 | 118,118 | 118,118 | q23.1-q24.1 |
| D16S363 | 51G1 | (GT)20 | 9(242-258) | 0.78 | 246,256 | 248,254 | q24.2-q24.3 |
| D16S392 | 305E9 | (GT)14+(TA)7+(T)10 | 9(113-151) | 0.78 | 119,121 | 115,119 | q24.2-q24.3 |
| D16S332 | 305D6 | (GT)13 | 3(122-128) | 0.53 | 122,122 | 122,126 | q24.2-q24.3 |
| D16S393 | 323H4 | (GT)25 | 14(130-158) | 0.87 | 152,152 | 140,150 | q24.2-q24.3 |
| D16S449 | 51A4 | (GT)21 | 9(183-201) | 0.85 | 193,197 | 195,195 | q24.2-q24.3 |
| D16S399 | MFD186# | (GT)16 | 10(192-210) | 0.71 | 198,198 | 198,206 | Chro. 1q |

# Provided by Dr. J.L.Weber (personal communication, 1991) was localized on human chromosome 1q by linkage analysis.

**Table 3-8.** The Categories of 32 STR Markers

| Repeat | | **Perfect Repeat** | Heterozygosity |
|---|---|---|---|
| D16S454 | 45G5 | (GT)22 | 0.75 |
| D16S524 | 40A7 | (GT)19 | 0.76 |
| D16S452 | 33A4 | (GT)24 | 0.68 |
| D16S288 | 7.1 | (GT)22 | 0.73 |
| D16S383 | 80B3 | (GT)15 | 0.45 |
| D16S531 | 8.15 | (GT)31 | 0.86 |
| D16S320 | 8.52 | (GT)24 | 0.86 |
| D16S389 | 10B3 | (GT)16 | 0.77 |
| D16S451 | 69F12 | (GT)21 | 0.84 |
| D16S347 | 12F8 | (GT)18 | 0.76 |
| D16S398 | MFD168 | (GT)22 | 0.85 |
| D16S397 | MFD98 | (GT)22 | 0.83 |
| D16S318 | 8.20 | (GT)22 | 0.54 |
| D16S522 | 8.21 | (GT)21 | 0.69 |
| D15S450 | 80H3 | (GT)21 | 0.52 |
| D16S395 | 33G11 | (GT)16 | 0.69 |
| D16S289 | 7.46 | (GT)21 | 0.77 |
| D16S40 | CRI.015 | (GT)15 | 0.11 |
| D16S363 | 51G1 | (GT)20 | 0.78 |
| D16S332 | 305D6 | (GT)13 | 0.53 |
| D16S393 | 323H4 | (GT)25 | 0.87 |
| D16S449 | 51A4 | (GT)21 | 0.85 |

Table 3-8. Continued

| Repeat | | Imperfect Repeat | Heterozygosity |
|---|---|---|---|
| D16S523 | 13H1 | $(GT)_{19}+(GT)_3$ | 0.68 |
| D16S453 | 301G12 | $(GT)_{11}+(GT)_4$ | 0.45 |
| D16S49 | CRI.0114 | $(GT)_6+(GT)_3+(GT)_{14}$ | 0.68 |
| D16S370 | 25H11 | $(GT)_7+(GT)_9$ | 0.30 |
| D16S317 | 7.9 | $(GT)_4+(GT)_4+(GT)_{13}$ | 0.16 |

| Repeat | | Compound Repeat | Heterozygosity |
|---|---|---|---|
| D16S525 | 308G7 | $(GT)_{10}(GA)_{21}(GT)_{21}$ | 0.91 |
| D16S319 | 7.14 | $(T)_6+(GT)_{17}+(T)_9$ | 0.52 |
| D16S359 | 26E3B | $(GT)_{12}+(GC)_4$ | 0.42 |
| D16S390 | 10F5 | $(GT)_{14}+(GC)_3+(GT)_7$ | 0.80 |
| D16S382 | 305E9 | $(GT)_{14}+(TA)_7+(T)_{10}$ | 0.78 |

**Table 3-9.** The Informativeness of 32 STR Markers

| Repeat | | **Repeat Units ≥ 16 (Max.Perfect Units)** | Heterozygosity |
|---|---|---|---|
| D16S525 | 308G7 | (GT)10(GA)21(GT)21 | 0.91 |
| D16S523 | 13H1 | (GT)19+(GT)3 | 0.68 |
| D16S454 | 45G5 | (GT)22 | 0.75 |
| D16S524 | 40A7 | (GT)19 | 0.76 |
| D16S452 | 33A4 | (GT)24 | 0.68 |
| D16S288 | 7.1 | (GT)22 | 0.73 |
| D16S319 | 7.14 | (T)6+(GT)17+(T)9 | 0.52 |
| D16S531 | 8.15 | (GT)31 | 0.86 |
| D16S320 | 8.52 | (GT)24 | 0.86 |
| D16S389 | 10B3 | (GT)16 | 0.77 |
| D16S451 | 69F12 | (GT)21 | 0.84 |
| D16S347 | 12F8 | (GT)18 | 0.76 |
| D16S398 | MFD168 | (GT)22 | 0.85 |
| D16S397 | MFD98 | (GT)22 | 0.83 |
| D16S318 | 8.20 | (GT)22 | 0.54 |
| D16S522 | 8.21 | (GT)21 | 0.69 |
| D15S450 | 80H3 | (GT)21 | 0.52 |
| D16S395 | 33G11 | (GT)16 | 0.69 |
| D16S289 | 7.46 | (GT)21 | 0.77 |
| D16S363 | 51G1 | (GT)20 | 0.78 |
| D16S393 | 323H4 | (GT)25 | 0.87 |
| D16S449 | 51A4 | (GT)21 | 0.85 |

**Table 3-9.** Continued

| Repeat | | **Repeat Units 9-15 (Max.Perfect Units)** | Heterozygosity |
|---|---|---|---|
| D16S453 | 301G12 | $(GT)_{11}+(GT)_4$ | 0.45 |
| D16S49 | CRI.0114 | $(GT)_6+(GT)_3+(GT)_{14}$ | 0.68 |
| D16S370 | 25H11 | $(GT)_7+(GT)_9$ | 0.30 |
| D16S383 | 80B3 | $(GT)_{15}$ | 0.45 |
| D16S359 | 26E3B | $(GT)_{12}+(GC)_4$ | 0.42 |
| D16S390 | 10F5 | $(GT)_{14}+(GC)_3+(GT)_7$ | 0.80 |
| D16S317 | 7.9 | $(GT)_4+(GT)_4+(GT)_{13}$ | 0.16 |
| D16S40 | CRI.015 | $(GT)_{15}$ | 0.11 |
| D16S332 | 305D6 | $(GT)_{13}$ | 0.53 |
| D16S382 | 305E9 | $(GT)_{14}+(TA)_7+(T)_{10}$ | 0.78 |

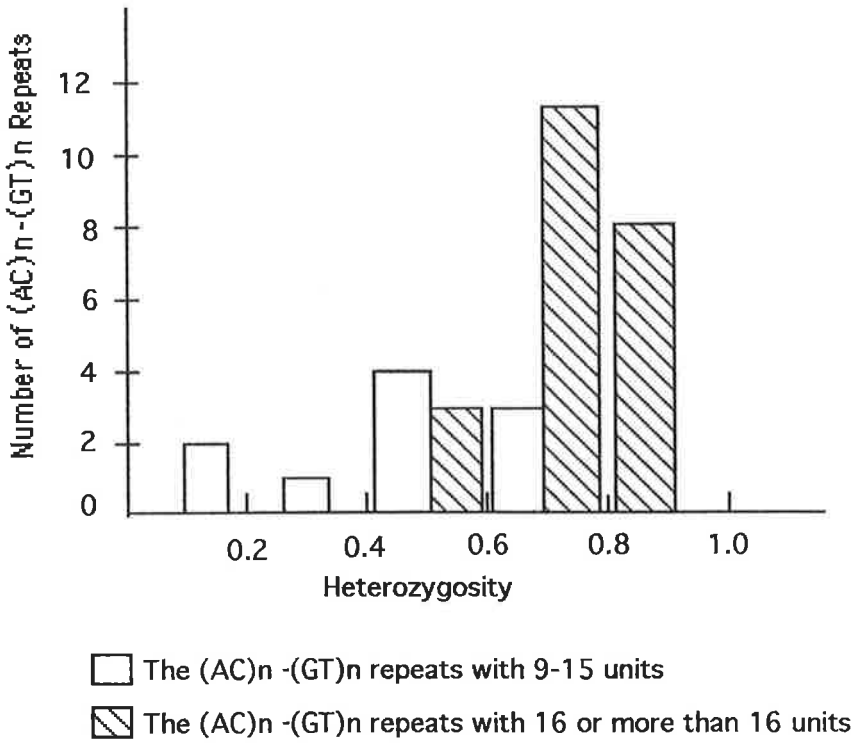



Figure 3-2. Repeat length and informativeness

increased with the numbers of perfect repeat units. The types of repeats (Table 3-8) did not show any relation to the increase in heterozygosity.

## 3.4. DISCUSSION

### 3.4.1. Isolation of STR Markers

Two approaches were used in this project to isolate simple tandem repeat markers for construction of a high-resolution genetic linkage map of human chromosome 16. The random isolation was initially used to isolate STR markers from the cosmid library of human chromosome 16 (Stallings et al. 1990). Of 25 (AC)n repeats, 10 (40%) were determined to be mouse (AC)n repeats. The large number of (AC)n mouse repeats isolated was an unexpected problem. This was greater than the level of mouse DNA contamination of the cosmid library, which was reported to be 6.5% (Stallings et al. 1990). (AC)n repeats are the most common STR repeats in the mouse genome but are the second most common repeats in the human genome, furthermore the (AC)n repeats in the mouse genome usually have more repeat units than those in human genome (Backmann and Weber 1992). Thus, the mouse (AC)n repeats had a higher chance of being isolated in this study.

Twelve (AC)n repeat markers were human repeats. However, in three instances (AC7.42 and AC1.18, AC7.46 and AC7.11, and AC7.55 and AC1.1)(Table 3-1), the identical (AC)n repeats were independently isolated. Since these three repeats contain

perfect (AC)n greater than 20, they were more likely to show strong signals during hybridization. The cosmid library from which the (AC)n repeats were isolated was fractionated to 200-500bp fragments, therefore these identical repeats appear not to be randomly isolated.

One repeat (*D16S317*, AC7.9) with a low heterozygosity (0.16, Table 3-7) could not be used as a genetic marker for construction of the genetic linkage map. AC7.22 detected two loci flanking *FRA16A* on chromosome 16 (Figure 4-3). AC7.59 detected more than one locus on chromosome 16, and also detected loci on chromosomes 12, 14, 18 and Y (Figure 4-1). The allele frequencies from loci of AC7.22 and AC7.59 could not be determined because of overlapping allele sizes. These two markers could not be used for map construction. In the end, there were only six markers (*D16S288, D16S319, D16S289, D16S318, D16S531* and *D16S320* ) (50%, 6/12) that could be used for construction of the genetic map of chromosome 16.

Weissenbach et al. (1992) used the random approach to isolate (AC)n repeat markers from the whole human genome. A total of 12,014 clones were sequenced and oligo primers were designed for 2,995 selected repeats. 1,547 repeats could be assigned to a chromosome. Finally, 813 (AC)n repeat markers (53%, 813/1,547) could be used for construction of a genetic map of the human genome. The results (50%, 6/12) of the random isolation of (AC)n repeats from the chromosome 16 cosmid library were similar to those of Weissenbach et al. (1992).

When 9 (AC)n repeats (6 of them were used for construction of genetic map) were physically mapped to 9 intervals (Figure 4-3), a targetted approach was adopted to isolate simple tandem repeats from 74 cosmid clones which had been previously mapped to specific regions of chromosome 16. Fifty-eight positive cosmid clones were subcloned into M13. Two (AC)n repeats isolated from two cosmids (302H7 and PKDcos5) were identical to an (AC)20 repeat (ACVK5, *D16S94* ) which had been isolated from a lambda clone VK5 in the PKD1 region (Hyland et al. 1990). A compound repeat AC305E9 isolated from cosmid 305E9 was identical to the repeat isolated from cosmid 311D9. These cosmids were subsequently shown to overlap. Since (AC)n repeats are distributed generally about one repeat every 30 kb in the human genome (Stallings et al. 1991), the isolation of (AC)n markers from the same physical region (especially from the same contig) may cause the isolation of identical repeats.

Of 41 repeats isolated from both approaches, 8 (AC)n repeats were discarded due to the detection of more than one locus, or failure of PCR amplification by adjacent repetitive sequences. Weber (1990a) reported that 7% of (AC)n repeats yield three or more alleles per individual. 9.8% (4/41) in this study is slightly higher. The (AC)n repeats adjacent to repetitive sequences, such as the *Alu* sequence, were also found by others to be difficult to amplify by PCR (Economou et al., 1990; Beckmann and Weber, 1992). Two (AC)n repeats (AC25H11 and

ACCRI.015) with low heterozygosities (0.30 and 0.11) were not be used as genetic markers.

A total of 21 (AC)n repeat markers isolated from 58 positive cosmid clones were used for construction of the genetic map of chromosome 16. The specific approach generated more genetic markers in known locations which were therefore useful for filling in the deficient regions and localizing disease genes (FMF, Batten disease).

3.4.2. Characterization of STR Markers

A total of 32 (AC)n repeat markers were characterized in this study. Twenty-seven markers (84%) had more than 5 alleles. Twenty-six markers (81%) had heterozygosities greater than 0.50 [ 22 of them (69%) had heterozygosities ≥ 0.68]. This confirms that the (AC)n repeats on chromosome 16 are highly polymorphic genetic markers useful for construction of a genetic linkage map of this chromosome, assigning and regional localizing disease genes to it.

The isolation of these simple tandem repeats from chromosome 16 revealed that the sequences differed from each other both in repeat types and in numbers of repeats. The informativeness of (AC)n repeat markers was found not to relate to the repeat types (Table 3-8), but to relate to the perfect repeat numbers (Table 3-9). Figure 3-2 shows that the informativeness of the (AC)n repeat markers generally increases with the increasing number of perfect repeat numbers. On average, the

dinucleotide repeats containing uninterrupted ≥ 16 units will predictably be highly informative.

J. L. Weber (1990a) assumed that there are about 7,000 (AC)n repeats with heterozygosity ≥ 0.70 in the human genome based on hybridization results and the distribution of cloned sequences. According to this assumption, and if (AC)n repeats are evenly distributed, and chromosome 16 is about 3% (Morton, 1991) of the human genome (3300 cM), therefore, chromosome 16 should have 210 (AC)n repeats with heterozygosities ≥ 0.70. Theoretically, these numbers of (AC)n repeat markers alone would yield a genetic linkage map of chromosome 16 with an average resolution of approximately 0.48 cM. There are 22 (AC)n repeat markers with heterozygosities ≥ 0.68 in this study (17 have a heterozygosity greater than 0.70).

# CHAPTER 4

## CONSTRUCTION OF PCR-BASED PHYSICAL AND GENETIC MAPS OF HUMAN CHROMOSOME 16

## 4.1. INTRODUCTION

Genetic linkage maps of human chromosomes are important tools for the localization of disease genes. One of the first five year goals of the Human Genome Project was the construction of genetic maps of all human chromosomes with highly polymorphic markers spaced an average of 2-5 centimorgans apart (Jordan, 1992). A number of genetic maps have been constructed for human chromosome 16 (Donis-Keller et al., 1987; Keith et al., 1990; Julier et al., 1990; NIH/CEPH Collaborative mapping group, 1992; Kozman et al., 1993). These maps were based mainly on RFLP markers which are less efficient for localization of disease genes and for refining linkage distance for positional cloning than highly polymorphic PCR based markers. Therefore, there was a need for more highly informative markers placed on high density cytogenetic-based physical and genetic linkage maps of chromosome 16.

The sex-averaged genetic length of human chromosome 16 has been estimated to be about 150 cM, which accounts for approximately 3% of the human genome, or an estimated 98 megabases of DNA (Morton, 1991). The expected number of genes on chromosome 16 is 1,500-3,000 (Mulley and Sutherland, 1993). 38 cloned genes and 15 cDNA sequences have been mapped onto chromosome 16 (Whitmore et al., 1993). Several disease genes have been localized on chromosome 16, including the genes for autosomal dominant adult polycystic kidney disease (PKD1) (Reeders et

al., 1988; Somlo et al., 1993), familial Mediterranean fever (FMF) (Shohat et al., 1992; Aksentijevich et al., 1993b), tuberous sclerosis (Kandt et al., 1993), Rubinstein-Taybi syndrome (Breuning et al., 1993), Batten disease (CLN3) (Mitchison et al., 1993; Yan et al., 1993) and Morquio A syndrome (Baker et al., 1993). Chromosome 16 contains four well characterized fragile sites *FRA16A* (16p13.11), *FRA16E* (16p12.1), *FRA16B* (16q21-q22.1) and *FRA16D* (16q23.1) (Sutherland, 1993).

Characteristics of the markers used to build linkage maps which can be improved are informativeness, abundance and ease of use. The STR markers have provided the means to make major improvements in these three characteristics. The STR markers are frequently highly polymorphic, of wide distribution in eukaryotic genomes, and can be easily analysed using PCR techniques (Weber et al., 1989; Litt et al., 1989; Tautz, 1989). These characteristics make them ideal genetic markers for construction of a high-resolution genetic linkage map of human chromosome 16.

## 4.2. MATERIALS AND METHODS

### 4.2.1. Physical Mapping of STR Markers

#### 4.2.1.1. Panel of Human Chromosome 16 Somatic Cell Hybrids (constructed by Dr. D. F. Callen)

The panel of chromosome 16 somatic cell hybrids was constructed by fusing human cell lines containing chromosome 16 translocations and deletions with the mouse line A9 and selecting for the human APRT gene located at 16q24 (Callen, 1986; Callen et al., 1990). The human cell lines were derived from patients identified in routine clinical cytogenetic investigations (Callen, 1986; Callen et al., 1990; Callen et al., 1992). This panel of 67 hybrids, together with four fragile sites, have the potential to divide chromosome 16 into 66 interval regions (Callen et al., 1992; Shen et al., submitted). The hybrid cell lines are depicted in Figure 4-3. This panel allowed the physical mapping of gene probes, anonymous DNA probes (Sequence Tagged Sites, STSs) and highly polymorphic DNA markers [ STRs (simple tandem repeats, microsatellites), variable number of tandem repeats (VNTRs, minisatellites) and RFLPs (restriction fragment length polymorphisms)].

#### 4.2.1.2. Physical Mapping of STR Markers by PCR

The physical location of each STR marker was determined by PCR analysis (Richards et al., 1991; Thompson et al., 1992) using DNAs from the human

chromosome 16 somatic cell hybrid panel as templates. PCR products were resolved on 1.4 or 2.0% agarose gels and visualized by ethidium bromide staining. This allowed mapping of each STR marker to a specific breakpoint interval by virtue of the presence or absence of a PCR product. The PCR for hybrid DNAs with breakpoints that flanked this interval helped to confirm the physical assignment. Some STRs which were difficult to resolve on agarose gels were amplified by incorporating [$\alpha$-$^{32}$P] dCTP into PCR products. Polyacrylamide gels were then used for resolving the presence or absence of the correct size PCR product.

NIGMS human-rodent somatic cell hybrid mapping panel 2 (Drwinga et al., 1993) was used for mapping some of the STRs which detected more than one locus in the human genome.

### 4.2.2. Genotyping of STR Markers

#### 4.2.2.1. CEPH Reference Family Panel

Families with large sibships, living parents and grandparents are especially informative for linkage mapping. From 100 families available from various sources, selected not for the presence of disease but for large sibship size, an initial group of 40 families were defined for the CEPH (Centre d'Etude du Polymorphisme Humain) reference family panel by a group of collaborating investigators (Dausset et al., 1990). These 40 families contain 121 grandparents (29 families with all four

grandparents), 40 pairs of parents and 334 offspring. No family has less than 6 offspring, and the 8 largest families (102, 884, 1331, 1332, 1347, 1362, 1413, 1416) have 10 or more offspring.

Of 36 STR markers genotyped through the CEPH family reference panel by the candidate (Table 4-2), 20 markers were genotyped in 40 families and a further 16 markers were genotyped in the same 8 families as those genotyped by Weissenbach et al.(1992).

4.2.2.2. Genotyping of STR Markers by PCR

Genotyping was carried out using the primer pair of each marker to amplify the DNA (100ng) of each individual from the CEPH reference families. The PCR reactions and amplification conditions used were as described in the PCR method section of chapter two. 18 STR markers were amplified using multiplex reactions in which two STR repeats could be amplified. The PCR samples were loaded on 3.5-5% polyacrylamide gels and electrophoresed. Each gel was loaded with a size marker, which was a PCR product obtained using the subcloned and sequenced M13 as template (Figure 4-4). The alleles were resolved on the gels after exposure to X-ray film. The genotype of each individual was determined by the allele segregations based on the parental genotypes.

The computer program CHROMPIC (an option of CRI-MAP) was used to detect double recombinants which could be possible genotyping errors. The autoradiographs of potential genotyping errors were checked and if necessary the original DNA samples were regenotyped (some samples were regenotyped twice or more).

### 4.2.3. Genetic Linkage Analysis

(performed by H.M. Kozman)

Linkage analysis was carried out using a Sun SPARC station IPC. Marker order and map distances were determined by likelihood analysis using the computer program CRI-MAP (Version 2.4) (Lander and Green, 1987). Using the BUILD option, the two loci with the greatest number of informative meioses, *D16S85* and *D16S320*, were used as the foundation of the map. The next most informative locus was then selected and added to the map if the interval support (Keats et al., 1991) was at least 3 (odds of 1000:1 in favour of one position over any other position on the map). This process was repeated until no further locus could be placed uniquely. This generated a framework genetic map (Keats et al., 1991).

The ALL option was used to order the remaining loci. They were added to the map by selecting the position with the highest likelihood. The FLIPS option was used to obtain odds against inversion of adjacent loci to provide a measure of

support for the placement of loci. This generated a comprehensive genetic linkage map (Keats et al., 1991) calculated using sex-averaged map distances between loci. Some loci placed with odds of less than 1000:1 by linkage could be ordered by physical mapping. The data were checked for errors using the CHROMPIC option of CRI-MAP. Once the genetic database had been corrected, the sex-averaged map was reconstructed and the sex-specific distances determined. Recombination frequency was converted to cM (centiMorgans) using the Kosambi mapping function.

## 4.3. RESULTS

### 4.3.1. Physical Mapping of STR Markers

#### 4.3.1.1. Mapping of STR Markers from the Random Isolation

9 STR markers (except 3 identical repeats) isolated from the random isolation were mapped to 9 breakpoint intervals (Figure 4-3) of chromosome 16 using the human-mouse somatic cell hybrid panel. 16AC7.22 detected two loci (16AC7.22A and B) which flank the *FRA16A* region on chromosome band 16p13.11(Figure 4-3). 16AC7.59 detected at least two loci with four alleles (Figure 4-1) on chromosome 16. The primers of this repeat also amplified repeat alleles on human chromosomes 12, 14, 18 and Y using the NIGMS human-rodent somatic cell hybrid mapping panel 2 (Figure 4-1). After checking the GenBank Database, the forward

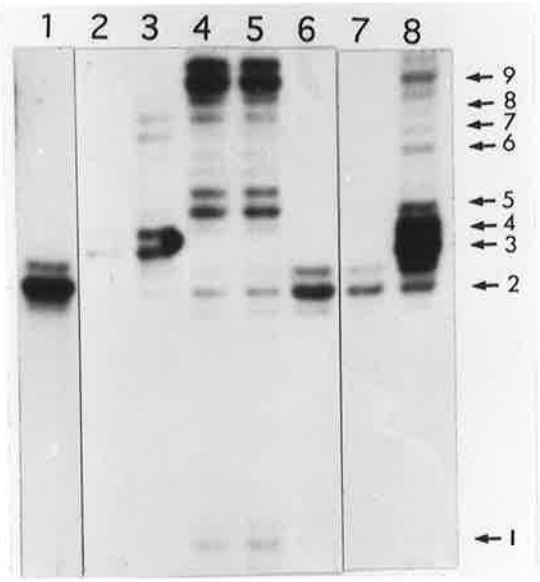



**Figure 4-1** Mapping of 16AC7.59

Lane 1. M13 clone had allele 2; Lane 2. Hybrid PK-89-12 which contains chromosome 12 as the only human chromosome had alleles 3 and 6 (faint); Lane 3. Hybrid HDm-5 which contains chromosome 14 as the only human chromosome had alleles 2 (faint), 3, 4, 7 and 8; Lanes 4. and 5. Two DNA preparation batches of CY18 which contain chromosome 16 as the only human chromosome had alleles 1, 2, 5 and 9; Lane 6. Hybrid E1B3 which contains chromosome 18 as the only human chromosome had allele 2; Lane 7. Hybrid 7631 which contains the Y chromosome as the only human chromosome had allele 2; Lane 8. Human genomic DNA had alleles 2, 3, 4, 6, and 9.

and reverse oligo primers were found to be part of the *Alu* sequence and had 70-100% homology to the intron sequences of more than ten genes and cDNAs. 16AC7.59 could not be mapped on chromosome 16 using the hybrid panel because the chromosome 16 hybrid panel contains some other human chromosomes, such as 12, 14, 18 and so on. 16AC8.20 (*D16S318* ) was remapped because of the discrepancy between physical and genetic localizations. The physical location of 16AC8.20 was between CY130(D) and CY4 (Table 4-1)(Figure 4-3).

4.3.1.2. Remapping of STR Markers from the Specific Isolation

8 STR markers isolated from specific regions of chromosome 16 were remapped because of the discrepancy between physical and genetic localizations, or the original locations of cosmid clones were not confidently determined. The locations of these 8 STR markers on the physical map were reevaluated and corrected (Table 4-1 )(Figure 4-3).

16AC39C8 was isolated from cosmid 39C8. The primer pair synthesized from the M13 subclone sequence produced three or four alleles in CEPH parent DNAs. In order to clarify whether these alleles existed on chromosome 16, the NIGMS human-rodent somatic cell hybrid mapping panel 2 was subsequently used to map them. The result showed this repeat also existed on human chromosomes 12 and 14 (Figure 4-2). By checking of the GenBank Database, the forward and reverse primer sequences were found to be part of the *Alu* sequence, and showed 80-100%

**Table 4-1.** Remapping of 9 STR Markers

| Marker | | Present Location* | Previous Location |
|---|---|---|---|
| D16S525 | 16AC308G7 | CY14-CY200 | PKD1 region |
| D16S523 | 16AC13H1 | 23HA,CY190,CY186--CY177,CY194,CY182 | CY163-CY11 |
| D16S49 | 16ACCRI.0114 | CY19-CY185 | CY180(D)-CY19 |
| D16S452 | 16AC33A4 | CY11-CY180(P) | CY180(P)-CY175 |
| D16S370 | 16AC25H11 | CY11-CY180(P) | CY180(P)-CY175 |
| D16S389 | 16AC10B3 | CY130(P)-CY122 | CY115-CY107(D) |
| D16S347 | 16AC12F8 | CY130(D)-CY4 | CY126-CY130(P) |
| D16S318 | 16AC8.20 | CY130(D)-CY4 | CY4-CY143 |
| D16S363 | 16AC51G1** | CY100,CY120-CY18A(P2) | CY18A(P2)-CY112 |

* See figure 4-3.
** This marker failed to amplify a satisfactory PCR product using hybrid cell lines. It was physically mapped by S. Apostolou using Southern blot analysis.

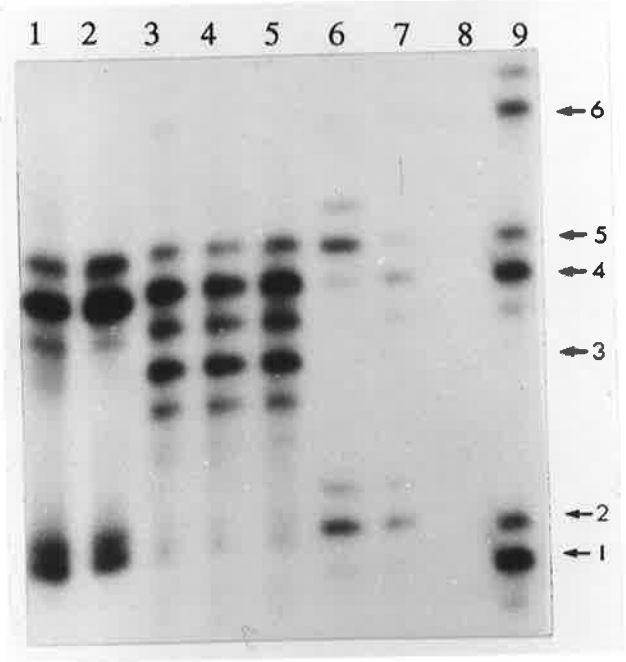



**Figure 4-2.** Mapping of 16AC39C8

Lane 1. M13 clone had allele 4; Lane 2. Cosmid clone had allele 4; Lanes 3, 4 and 5. Three DNA preparation batches of hybrid line CY 18 which contains chromosome 16 as the only human chromosome had alleles 3 and 4; Lane 6. Hybrid HDm-5 which contains chromosome 14 as the only human chromosome had alleles 2 and 5; Lane 7. Hybrid PK-89-12 which contains chromosome 12 as the only human chromosome had alleles 2 and 4; Lane 8. Hybrid J1C14 which contains chromosome 11 as the only human chromosome did not detect any allele ; Lane 9. Human genomic DNA had alleles 1, 4 and 6.

homology to the intron sequences of more than ten genes and cDNAs. The cosmid clone 39C8 from which 16AC39C8 was isolated was mapped to 16p12 by *in situ* hybridization using cosmid DNA as a probe (by H. Eyre). 16AC1E5 also produced more than two alleles on chromosome 16. These alleles were localized to the breakpoint interval between CY11 and CY163 (Figure 4-3).

4.3.1.3. Mapping of STR Markers from Other Laboratories

Twenty-four STR markers ( *D16S400-423* ), which were characterized and genetically mapped by Weissenbach et al.( 1992) were physically mapped to 14 breakpoint intervals of chromosome 16 using the hybrid cell panel (Figure 4-3). Two markers ( *D16S400* and *D16S409* ) were not mapped due to failure of PCR amplification.

Two STR markers mfd98 (*D16S397*) and mfd168 (*D16S398*) from Dr. J.L.Weber (personal communication, 1991) were mapped to one breakpoint interval between CY130(D) and CY4 (Figure 4-3). One (AC)n repeat marker (*D16S399*, mfd186) also from Dr. J.L. Weber was difficult to physically map. After linkage analysis (by genotyping through the CEPH families) the location of this marker was found to be on human chromosome 1q.

Three STR markers (*D16S310*, *D16S312* and *D16S313* ) from Hudson et al. (1992) were mapped to 3 breakpoint intervals of chromosome 16 (Figure 4-3).

*D16S312* detected two loci between breakpoint intervals CY185 and CY19 on chromosome 16p13.11 (Figure 4-3), which was localized by using the NIGMS human-rodent somatic cell hybrid mapping panel 2 and the chromosome 16 somatic cell hybrid panel.

4.3.1.4. The Physical Map of STR Markers of Chromosome 16

A total of **32** STR markers isolated from the two approaches were localized to 23 physical intervals (Figure 4-3) of a high resolution cytogenetic map of chromosome 16 (Callen et al., 1992; Shen et al., Submitted).

In addition, **33** STR markers from other laboratories [*D16S401-423* (Weissenbach et al., 1992); *D16S397-398* (J.L.Weber, personal communication 1991) and *D16S310, D16S312* and *D16S313* (Hudson et al., 1992); *D16S283* (Harris et al. 1991); SPN (Regaev et al. 1992); *D16S285* (Konradi et al. 1991); *D16S261, D16S265* and *D16S266* (Weber et al. 1990b)], **20** STR markers from colleagues [*D16S291-294* and *D16S298-308* (Thompson et al.; 1992); *D16S287* (Phillips et al., 1991a); *D16S164, 168* and*186* (Phillips et al., 1991b); *D16S94* (Aksentijevich et al., 1993b) and *D16S79A, D16S67, D16S295-297* (Callen et al. 1993)] mapped in the candidate's laboratory are shown on the high resolution cytogenetic map of chromosome 16 (Figure 4-3). The 67 human-mouse somatic cell hybrid lines, together with the four fragile sites, divided chromosome 16 into 66 breakpoint interval regions (Callen et al., 1992; Shen et al., submitted), and allowed

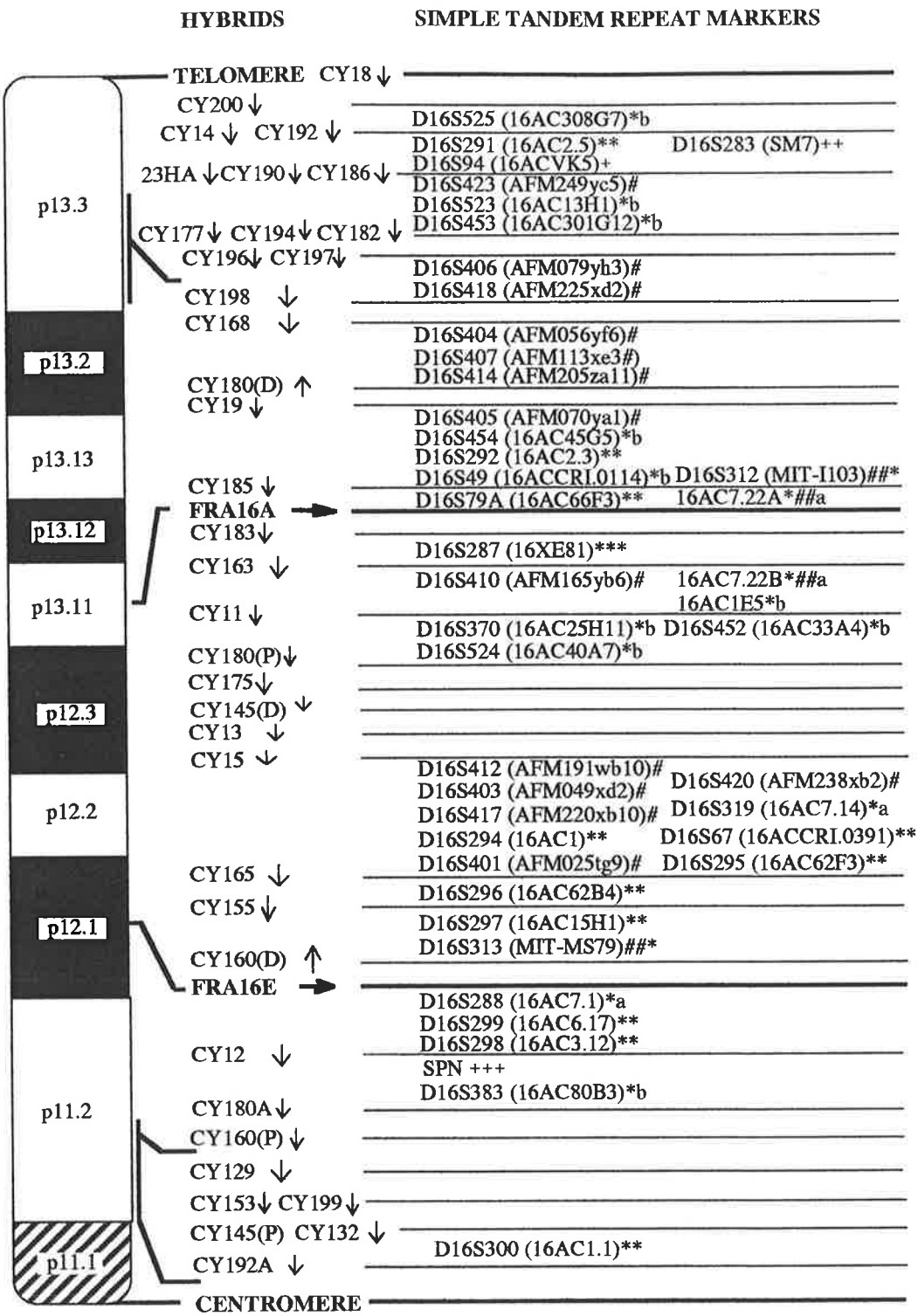
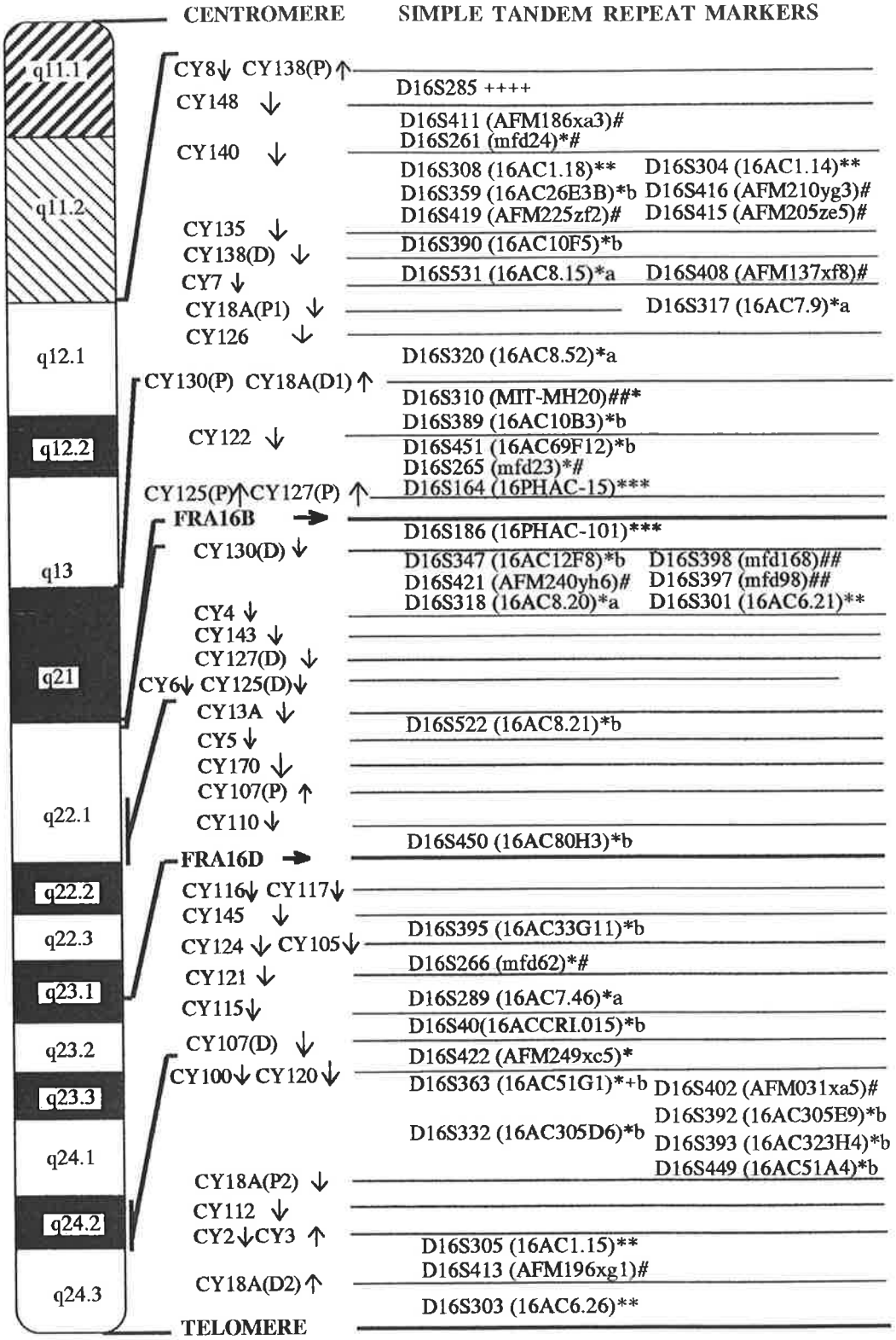



**Figure 4-3.** The Physical Map of 85 STR Markers of Chromosome 16

The portion of chromosome 16 present in each human/mouse hybrid cell line is delineated by a horizontal line with an arrow indicating the direction of the retained portion of chromosome 16 from the breakpoint. The breakpoint of the fragile sites is indicated by horizotal arrows. The 32 STR markers isolated by the candidate from random isolation and specific isolation are indicated by a and b, respectively.




* isolated and mapped by the candidate.
\# from Weissenbach et al. (1992), mapped by the candidate.
\## from J.L.Weber (personal communication,1991), mapped by the candidate.
*## isolated by the candidate and mapped by the candidate and H.A.Phillips.
*+ isolated by the candidate, mapped by S. Apostolou using Southern analysis.
##* from Hudson et al. (1992), mapped by the candidate and L-Z.Chen.
** isolated and mapped by A.Thompson.
*** isolated and mapped by H.A.Phillips.
*# from Weber et al. (1991), Mapped by H.A.Phillips.
\+ isolated by K.Holman from a lambda clone VK5 in the PKD1 region.
++ from Harris et al. (1991), mapped in the PKD region..
+++ from Rogaev et al. (1992), mapped by S. Whitmore.
++++ from Konradi et al (1991), mapped by A. Thompson.

A.Thompson, H.A.Phillips, K.Holman, L-Z.Chen, S.Whitmore and S.Apostolou were members of the Dept.of Cytogenetics and Molecular Genetics, Women's and Children's Hospital.

a total of **85** STR markers to be mapped to 37 breakpoint intervals of a high-resolution cytogenetic map of chromosome 16 (Figure 4-3).

### 4.3.2. Genotyping of STR Markers

#### 4.3.2.1. Genotyping of STR Markers on CEPH Reference Family Panel

A total of 8,231 genotypes were determined for the **36** STR markers [ **27** STR markers from the random and specific isolations, **5** markers (*D16S94, D16S291, D16S300, D16S301* and *D16S308* ) generated by A. Thompson (Thompson et al., 1992) and K. Holman (Aksentijevich et al., 1993b) and **4** from other laboratories (*D16S397* and *D16S398* from Dr.J.L.Weber, personal communication, 1991; *D16S310* and *D16S313* from Hudson et al., 1992) through the CEPH reference family panel. 20 markers were genotyped on 40 families and 16 were genotyped in the 8 same families (102, 884, 1331, 1332, 1347, 1362, 1413, 1416) as used by Weissenbach et al. (1992) (Table 4-2).

Each STR marker was observed to obey Mendelian codominant inheritance in CEPH families (Figure 4-4). The genotype of each individual was determined by the allele segregation based on the parental genotypes. An example of genotyping 16AC1.1 (*D16S300*) and 16AC6.21 (*D16S301*) on CEPH family 66 is shown in Figure 4-4.

**Table 4-2 .** 36 STR Markers Genotyped on CEPH Families

| 40 CEPH Families | | 8 CEPH Families | |
|---|---|---|---|
| D16S291 | 16AC2.5 | D16S525 | 16AC308G7 |
| D16S94 | 16ACVK5 | D16S523 | 16AC13H1 |
| D16S453 | 16AC301G12 | D16S524 | 16AC40A7 |
| D16S49 | 16ACCRI.0114 | D16S452 | 16AC33A4 |
| D16S319 | 16AC7.14** | D16S531 | 16AC8.15 |
| D16S288 | 16AC7.1** | D16S347 | 16AC12F8 |
| D16S313 | MIT-MS79 | D16S389 | 16AC10B3 |
| D16S359 | 16AC26E3B | D16S397 | 16ACMFD98 |
| D16S390 | 16AC10F5* | D16S451 | 16AC69F12 |
| D16S398 | 16ACMFD168 | D16S318 | 16AC8.20 |
| D16S301 | 16AC6.21 | D16S363 | 16AC51G1 |
| D16S450 | 16AC80H3 | D16S308 | 16AC1.18 |
| D16S522 | 16AC8.21* | D16S392 | 16AC305E9 |
| D16S395 | 16AC33G11* | D16S393 | 16AC323H4 |
| D16S289 | 16AC7.46 | D16S449 | 16AC51A4 |
| D16S310 | MIT-MH20* | D16S454 | 16AC45G5 |
| D16S320 | 16AC8.52 | | |
| D16S383 | 16AC80B3*** | | |
| D16S332 | 16AC305D6 | | |
| D16S300 | 16AC1.1 | | |
| 20 Markers | | 16 Markers | |

* Genotyped by the candidate and L-Z. Chen.
** Genotyped by A. Thompson.
*** Genotyped by H.A. Phillips.

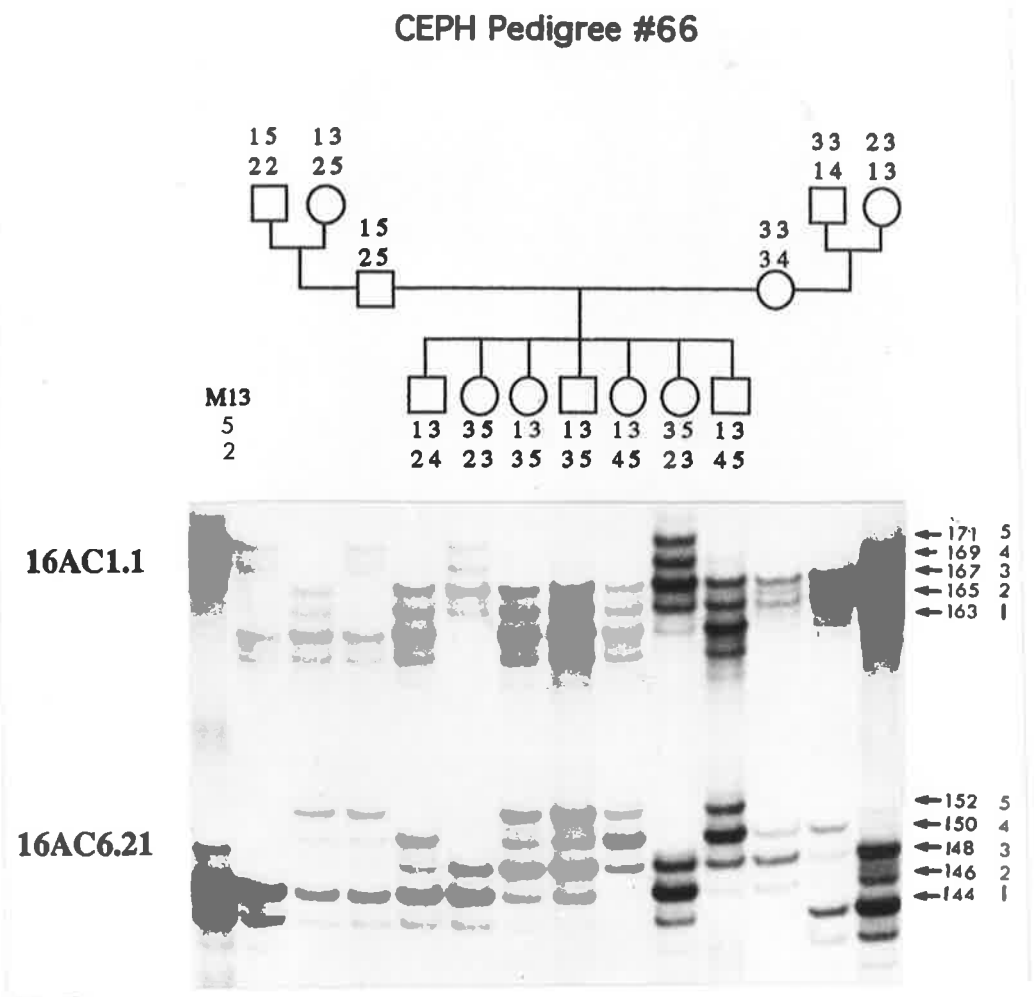



**Figure 4-4.** Genotyping of 16AC1.1 and 16AC6.21 through CEPH Family 66

A total of 16,462 parent-offspring transfers of alleles of 36 STR markers located on chromosome 16 were examined through the genotyping within 40 CEPH reference families. Five mutation events (Table 4-3) (Figure 4-5) which did not show Mendelian codominant inheritance were detected in these markers using DNAs from transformed lymphoblastoid cell lines. The mutation rate for the 36 chromosome 16 STR markers in this study was $3.0 \times 10^{-4}$ per locus per gamete per generation. Of five mutation events, two involved the gain of a single repeat unit [2bp for *D16S393* (AC323H4) and 4bp for tetranucleotide repeat *D16S310* (MIT-MH20)], and *D16S289* (AC7.46) involved the loss of 3 repeat units (Table 4-3 and Figure 4-5).

Of the two mutation events, *D16S313* (MIT-MS79) from family 1331 and *D16S301* (16AC6.21) from family 1347, both showed non-paternal alleles transmitted to offspring (Table 4-3 and Figure 4-5). At the *D16S313* locus (MIT-MS79), the grandfather (GF) had alleles 143(bp) and 147(bp), neither of which were present in his daughter (M) who was homozygous for allele 145(bp), which was presumably inherited from the grandmother (no DNA available), but allele 147(bp) was present in 5 children (C2, C3, C8, C10 and C11). This can be interpreted as 1) the allele 147(bp) was lost in somatic cells during or after the establishment of the transformed lymphoblastoid cell line, or 2) the alleles 147(bp) or 143(bp) were mutated to allele 145(bp) by losing or increasing one repeat unit. At the *D16S301* locus (16AC6.21), child 8 (C8) had allele 148(bp) from the mother

**Table 4-3 .** The Mutation Events of the Simple Tandem Repeats

| Marker | Sample No. | Genotype (bp) | Parent Genotype (bp) |
|---|---|---|---|
| 16AC7.46 | 884 C3 | 158, 168 | Father 164, 164<br>Mother 164, 168 |
| 16AC323H4 | 102 C12 | 148, 150 | Father 150, 152<br>Mother 146, 152 |
| MIT-MH20 | 1340 C13 | 166, 170 | Father 166, 166<br>Mother 158, 166 |
| MIT-MS79 | 1331 M | 145, 145 | G.Father 143,147<br>G.Mother ( No DNA) |
| 16AC6.21 | 1347 C8 | 148, 148 | Father 144, 144<br>Mother 144, 148 |

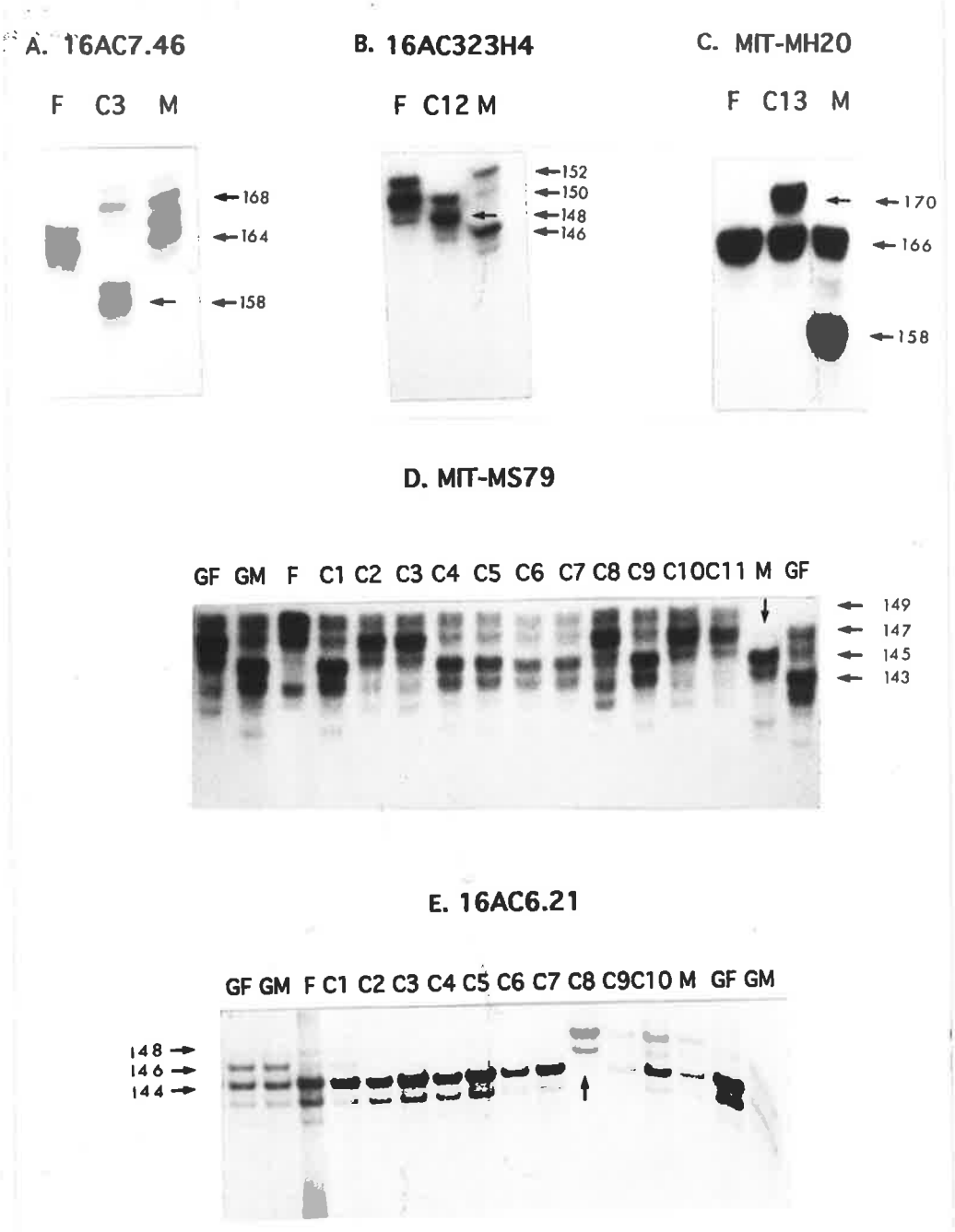



Figure 4-5. Mutations

A, B and C. Arrows indicate the mutation alleles.

D. The arrow indicates the individual (M) did not have her father's allele (147bp) which was present in her 5 children.

E. The arrow indicates the individual (C8) did not inherited the allele (144bp) from his father.

or 143(bp) were mutated to allele 145(bp) by losing or increasing one repeat unit. At the *D16S301* locus (16AC6.21), child 8 (C8) had allele 148(bp) from the mother (M) but no allele 144(bp) from the father (F) who was homozygous for allele 144(bp). This could be due to 1) somatic loss of allele 144(bp) in child 8, 2) allele 144(bp) mutated to allele 148(bp) by increasing 2 repeat units, or 3) the mutation at the primer site. After genotyping 55 STR markers, no other abnormal transmissions were found in these two families.

These observed mutation events could not be verified as germline mutations by excluding the possibility of somatic mutation in cell lines, because DNAs from untransformed cells were not available in this study. Hence, these mutations were eliminated in the linkage analysis.

The CHROMPIC option of CRI-MAP was used to detect double recombinants. When these were detected, the original data were reexamined and those DNA samples were regenotyped. False double recombination events were eliminated by this procedure. Thirteen double recombinants within a 15 cM region (9 of them within 10 cM) still remained (Table 4-4). These double recombinants were included in the linkage analysis.

**Table 4-4.** Double Recombination Events in Genotyping of 55 STR Markers

| Marker | | Individual | Genetic Distance (cM) | Maternal or Paternal Origin |
|---|---|---|---|---|
| D16S291 | 16AC2.5 | 1375-08* | 5.4 | M |
| D16S299 | 16AC6.17 | 1345-09 | 0.6 | P |
| D16S300 | 16AC1.1 | 1350-07* | 3.6 | P |
| D16S285 | | 21-04 | 2.7 | M |
| D16S261 | MFD24 | 37-04 | 2.5 | P |
| D16S359 | 16AC26E3B | 1408-04* | 2.3 | P |
| D16S390 | 16AC10F5 | 1344-07* | 15 | P and M |
| D16S310 | MIT-MH20 | 104-08* | 11.1 | M |
| D16S265 | MFD23 | 1345-05 | 4.8 | P |
| D16S522 | 16AC8.21 | 102-12* | 6.1 | P |
| D16S522 | 16AC8.21 | 1349-03* | 10.8 | P |
| D16S450 | 16AC80H3 | 1334-09* | 9.5 | M |

* double recombinant events in the genotyping of 36 STR markers.

### 4.3.3. Genetic Linkage Analysis

A total of 81 polymorphic markers [79 STR markers, 1 VNTR (D16S85 ) and 1 RFLP (D16S44)] were used to develop the PCR-based genetic linkage map of human chromosome 16. Of 81 markers, 27 STR markers (Table 4-5) were isolated by the candidate, the sources of the rest of the markers are listed in Table 4-5.

#### 4.3.3.1. The Framework Genetic Map

The framework genetic map contains 48 markers (Figure 4-6), one of which is the VNTR marker (3'HVR, *D16S85* ) (Kozman et al. 1993). *D16S85* (heterozygosity: 0.97) and *D16S320* (heterozygosity: 0.86) are two of the most informative markers and were used to initiate construction of the framework map. Each of the 48 markers can be placed with likelihood odds of 1000:1 or greater in favour of one position over any other position on the map. The order of the framework loci established by linkage was compatible with their physical order independently determined on the cytogenetic map (Figure 4-3 and 4-6). The median distance and the averaged distance between markers in the sex-averaged map is 2.7 cM and 3.2 cM, respectively. The length of the sex-averaged map is 152.1 cM. The largest gap between loci is 15.9 cM.

**Table 4-5.** The Sources of Chromosome 16 Simple Tandem Repeats

| Locus | Clone | Polymorphism Type | Reference |
|---|---|---|---|
| D16S525 | 308G7**a | $(GT)_{10}(GA)_{21}(GT)_{21}$ | Shen et al. 1994a |
| D16S291 | AC2.5* | $(GT)_{25}$ | Thompson et al. 1992 |
| D16S283 | SM7 | $(AC)_{19}$ | Harris et al. 1991 |
| D16S94 | VK5B** | $(GT)_{20}$ | Aksentijevich et al. 1993b |
| D16S523 | 13H1**a | $(GT)_{19}+(GT)_{3}$ | Shen et al. 1993d |
| D16S423 | AFM249yc5# | $(AC)_{17}$ | Weissenbach et al. 1992 |
| D16S453 | 301G12**a | $(GT)_{11}+(GT)_{4}$ | Shen et al. 1993c |
| D16S406 | AFM079yh3# | $(AC)_{22}+(AC)_{4}$ | Weissenbach et al. 1992 |
| D16S418 | AFM225xd2# | $(AC)_{24}$ | Weissenbach et al. 1992 |
| D16S404 | AFM056yf6# | $(CT)_{14}+(AC)_{13}$ | Weissenbach et al. 1992 |
| D16S407 | AFM113xe3# | $(AC)_{20}$ | Weissenbach et al. 1992 |
| D16S414 | AFM205za11# | $(AC)_{19}$ | Weissenbach et al. 1992 |
| D16S405 | AFM070ya1# | $(AC)_{25}$ | Weissenbach et al. 1992 |
| D16S454 | 45G5**a | $(GT)_{22}$ | Shen et al. 1993c |
| D16S292 | AC2.3* | $(GCT)_{7}(GT)_{18}$ | Thompson et al. 1992 |
| D16S49 | CRI.0114**a | $(GT)_{6}+(GT)_{3}+(GT)_{14}$ | Shen et al. 1993c |
| D16S79A | 66F3** | $(GT)_{20}$ | Phillips et al. 1993 |
| D16S287 | 16EX81** | $(GT)_{23}$ | Phillips et al. 1991a |
| D16S410 | AFM165yb6# | $(AC)_{13}$ | Weissenbach et al. 1992 |
| D16S452 | 33A4**a | $(GT)_{24}$ | Shen et al. 1993d |
| D16S524 | 40A7** | $(GT)_{19}$ | Shen et al. 1993d |
| D16S412 | AFM191wb10# | $(AC)_{4}+(AC)_{21}$ | Weissenbach et al. 1992 |

* randomly isolated from chromosome 16 cosmid library. *a isolated by the candidate.
** isolated from cosmid clones mapped in the specific regions. **a isolated by the candidate.
*** provided by Dr. J.L.Weber (personal communication, 1991) were characterized, physically mapped and genotyped by the candidate.
# physically mapped by the candidate.
## physically mapped and genotyped by the candidate and L-Z Chen.

**Table 4-5.** Continued (1)

| Locus | Clone | Polymorphism Type | Reference |
|---|---|---|---|
| D16S403 | AFM049xd2# | $(AC)_{23}$ | Weissenbach et al. 1992 |
| D16S417 | AFM220xb10# | $(AC)_{15}+(AC)_{2}$ | Weissenbach et al. 1992 |
| D16S294 | AC1* | (GT)n | Thompson et al. 1992 |
| D16S420 | AFM238xb2# | $(AC)_{14}$ | Weissenbach et al. 1992 |
| D16S401 | AFM025tg9# | $(AC)_{19}$ | Weissenbach et al. 1992 |
| D16S319 | AC7.14*a | $(GT)_{17}+(T)_{8}$ | Shen et al. 1992 |
| D16S67 | CRI.0391** | $(GT)_{23}$ | Callen et al. 1993 |
| D16S295 | 62F3** | $(TG)_{16}$ | Callen et al. 1993 |
| D16S296 | 62B4** | $(GT)_{16}$ | Callen et al. 1993 |
| D16S297 | 15H1H** | $(GT)_{7}+(TG)_{15}$ | Callen et al. 1993 |
| D16S313 | MIT-MS79## | $(CA)_{20}$ | Hudson et al 1992 |
| D16S288 | AC7.1*a | $(GT)_{22}$ | Shen et al. 1991 |
| D16S299 | AC6.17* | $(GT)_{19}$ | Thompson et al. 1992 |
| D16S298 | AC3.12* | $(GT)_{8}+(GT)_{24}$ | Thompson et al. 1992 |
| SPN## | | $(TC)_{11}(AC)_{13}(GC)_{12}$ | Rogaev et al. 1992 |
| D16S383 | 80B3**a | $(GT)_{15}$ | Shen et al. 1993d |
| D16S300 | AC1.1* | $(GA)_{12}(GT)_{24}(GA)_{5}$ | Weissenbach et al. 1992 |
| D16S285 | | (GT)n | Konradi et al. 1991 |
| D16S411 | AFM186xa3# | $(AC)_{16}$ | Weissenbach et al. 1992 |
| D16S261 | MFD24 | $(AC)_{7}+(AC)_{14}$ | Weber et al. 1990b |
| D16S308 | AC1.18* | $(GT)_{17}$ | Thompson et al. 1992 |
| D16S304 | AC1.14* | $(GT)_{20}$ | Thompson et al. 1992 |
| D16S359 | 26E3B**a | $(GT)_{12}+(GC)_{4}$ | Shen et al. 1994a |
| D16S416 | AFM210yg3# | $(AC)_{3}+(AC)_{14}$ | Weissenbach et al. 1992 |
| D16S415 | AFM205ze5# | $(AC)_{22}$ | Weissenbach et al. 1992 |
| D16S419 | AFM225zf2# | $(AC)_{15}(AG)_{12}$ | Weissenbach et al. 1992 |

**Table 4-5.** Continued (2)

| Locus | Clone | Polymorphism Type | Reference |
|---|---|---|---|
| 16S390 | 10F5**a | $(GT)_{14}+(GC)_3+(GT)_7$ | Shen et al. 1993b |
| D16S408 | AFM137xf8# | $(AG)_6+(AC)_{17}$ | Weissenbach et al. 1992 |
| D16S531 | AC8.15*a | $(GT)_{31}$ | Shen et al. 1994a |
| D16S320 | AC8.52*a | $(GT)_{24}$ | Shen et al. 1992 |
| D16S310 | MIT-MH20## | $(ATAG)_{12}$ | Hudson et al. 1992 |
| D16S389 | 10B3**a | $(GT)_{16}$ | Shen et al. 1993b |
| D16S451 | 69F12**a | $(GT)_{21}$ | Shen et al. 1993b |
| D16S265 | MFD23 | $(AC)_{20}$ | Weber et al. 1990b |
| D16S164 | 16PHAC-15** | $(GT)_{11}+(GT)_4(GA)_9$ | Phillips et al. 1991b |
| D16S186 | 16PHAC-101** | $(GT)_{13}$ | Phillips et al. 1991b |
| D16S347 | 12F8**a | $(GT)_{18}$ | Shen et al. 1993b |
| D16S421 | AFM240yh6# | $(AC)_{18}$ | Weissenbach et al. 1992 |
| D16S398 | MFD168 | $(GT)_{22}$ | *** |
| D16S397 | MFD98 | $(GT)_{22}$ | *** |
| D16S301 | AC6.21* | $(GT)_{19}$ | Thompson et al. 1992 |
| D16S522 | AC8.21**a | $(GT)_{21}$ | Shen et al 1994a |
| D16S318 | AC8.20*a | $(GT)_{22}$ | Shen et al. 1992 |
| D16S450 | 80H3**a | $(GT)_{21}$ | Shen et al. 1993b |
| D16S395 | 33G11**a | $(GT)_{16}$ | Shen et al. 1993b |
| D16S266 | MFD62 | $(AC)_{21}$ | Weber et al. 1990b |
| D16S289 | AC7.46*a | $(GT)_{21}$ | Shen et al. 1992 |
| D16S422 | AFM249xc5# | $(AC)_{19}+(AC)_7$ | Weissenbach et al. 1992 |
| D16S363 | 51G1**a | $(GT)_{20}$ | Shen et al. 1993a |
| D16S402 | AFM031xa5# | $(AC)_{23}$ | Weissenbach et al. 1992 |
| D16S392 | 305E9**a | $(GT)_{14}+(TA)_7+(T)_{10}$ | Shen et al. 1993a |
| D16S332 | 305D6**a | $(GT)_{13}$ | Shen et al. 1993a |

**Table 4-5.** Continued (3)

| Locus | Clone | Polymorphism Type | Reference |
|---|---|---|---|
| D16S393 | 323H4**a | $(GT)_{25}$ | Shen et al. 1993a |
| D16S449 | 51A4**a | $(GT)_{21}$ | Shen et al. 1993a |
| D16S305 | AC1.15* | $(GT)_{27}(AT)_{6}+(GT)_{6}$ | Thompson et al. 1992 |
| D16S413 | AFM196xg1# | $(A)_{15}+(AC)_{24}(AT)_{5}$ | Weissenbach et al. 1992 |
| D16S303 | AC6.26* | $(GT)_{9}+(GT)_{24}+(GT)_{7}$ | Thompson et al. 1992 |

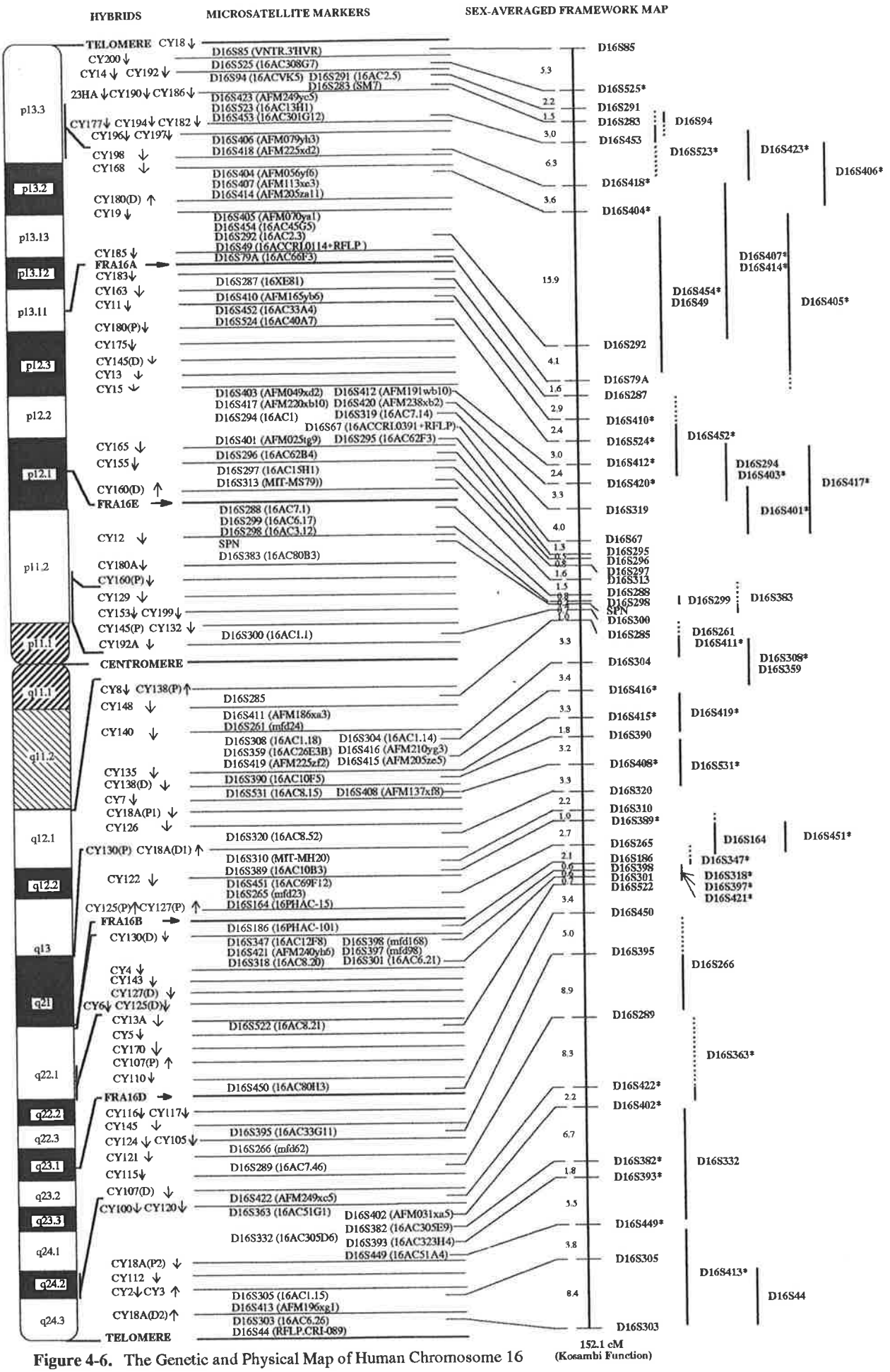



**Figure 4-6.** The Genetic and Physical Map of Human Chromosome 16

4.3.3.2. The Comprehensive Genetic Map

32 STR markers and one RFLP marker (*D16S44* ) (Kozman et al. 1993), in addition to the 48 framework markers, were incorporated into the comprehensive genetic map (Table 4-6). These 33 markers with an interval support of less than 3 (1000:1 odds) could not be uniquely ordered on the genetic map by linkage analysis (Figure 4-6). All of these markers could be anchored, however, by physical mapping using the chromosome 16 hybrid panel. These 33 markers were placed within 18 different breakpoint intervals (Figure 4-6). The average interlocus distance in the comprehensive map is 1.9 cM. The largest gap between loci is 13.6 cM.

The distances between the loci in the sex-specific maps are shown in Table 4-6. The male map is 126.8 cM long with an average interlocus distance of 1.6 cM, and the largest gap is 17.4 cM. The female map is 178.9 cM with an average interlocus distance of 2.2 cM, and the largest gap is 10.3 cM. The overall female to male ratio of recombination in the sex-specific maps is 1.4 : 1. On both distal regions of the chromosome 16 arms, the recombination rates on the male map are greater than the female map (2.9 fold between *D16S85* and *D16S406* on the distal short arm and 4.0 fold between *D16S449* and *D16S303* on the distal long arm). An excess of female over male recombination rates was seen in most genetic intervals throughout the remainder of this chromosome.

**Table 4-6.** Distances (in cM) between the loci in the comprehensive map, including sex-specific and sex-average distances and likelihood support for order

| Locus Interval | Male | Female | Sex-average | Odds(1:) |
|---|---|---|---|---|
| D16S85 | 8.3 | 2.6 | 5.2 | $10^{11}$ |
| D16S525 | 4.0 | 0.0 | 2.0 | $10^{5}$ |
| D16S291 | 1.2 | 0.8 | 1.1 | $10^{6}$ |
| D16S283 | 0.0 | 1.0 | 0.4 | $10^{2}$ |
| D16S94 | 2.3 | 0.6 | 1.6 | $10^{3}$ |
| D16S523 | 2.6 | 1.3 | 1.8 | $10^{2}$ |
| D16S423 | 0.0 | 0.0 | 0.0 | 1.0 |
| D16S453 | 8.4 | 3.0 | 5.8 | $10^{15}$ |
| D16S406 | 0.0 | 0.0 | 0.0 | 1.0 |
| D16S418 | 2.7 | 5.6 | 4.1 | $10^{13}$ |
| D16S404 | 0.0 | 0.0 | 0.0 | 1.0 |
| D16S407 | 0.0 | 2.6 | 1.5 | $10^{3}$ |
| D16S414 | 17.4 | 10.3 | 13.6 | $10^{21}$ |
| D16S405 | 0.0 | 0.0 | 0.0 | 1.0 |
| D16S454 | 0.0 | 0.0 | 0.0 | 1.0 |
| D16S292 | 0.8 | 0.0 | 0.4 | 10 |
| D16S49 | 4.0 | 4.1 | 4.0 | $10^{15}$ |
| D16S79A | 2.1 | 0.8 | 1.6 | $10^{11}$ |
| D16S287 | 1.2 | 5.5 | 3.0 | $10^{10}$ |
| D16S410 | 0.5 | 2.8 | 2.0 | $10^{2}$ |
| D16S452 | 0.8 | 0.0 | 0.0 | 1.0 |
| D16S524 | 1.2 | 6.5 | 3.9 | $10^{7}$ |
| D16S412 | 0.0 | 3.2 | 1.3 | 2.0 |
| D16S403 | 0.0 | 0.0 | 0.0 | 1.0 |
| D16S417 | 0.0 | 0.0 | 0.0 | 1.0 |
| D16S294 | 0.0 | 1.1 | 0.6 | $10^{3}$ |
| D16S420 | 0.0 | 5.3 | 2.4 | $10^{9}$ |
| D16S401 | 1.5 | 0.0 | 1.2 | $10^{2}$ |
| D16S319 | 2.5 | 5.3 | 3.7 | $10^{28}$ |
| D16S67 | 0.0 | 2.1 | 1.3 | $10^{8}$ |
| D16S295 | 1.9 | 0.0 | 0.5 | $10^{6}$ |
| D16S296 | 0.0 | 1.4 | 0.7 | $10^{14}$ |

**Table 4-6.** Continued (1)

| Locus Interval | Male | Female | Sex-average | Odds(1:) |
|---|---|---|---|---|
| D17S297 | 0.0 | 3.0 | 1.6 | $10^{24}$ |
| 16S313 | 0.3 | 2.6 | 1.4 | $10^{23}$ |
| D16S288 | 0.3 | 1.5 | 0.9 | $10^{11}$ |
| D16S299 | 0.3 | 0.0 | 0.3 | 1.6 |
| D16S298 | 0.0 | 0.9 | 0.3 | $10^{4}$ |
| SPN | 0.1 | 0.0 | 0.0 | 1.0 |
| D16S383 | 0.2 | 0.8 | 0.7 | $10^{5}$ |
| D16S300 | 0.0 | 1.9 | 1.0 | $10^{8}$ |
| D16S285 | 0.2 | 0.0 | 0.0 | 1.0 |
| D16S411 | 0.3 | 0.8 | 0.5 | $10^{3}$ |
| D16S261 | 0.3 | 7.3 | 3.9 | $10^{17}$ |
| D16S308 | 0.2 | 0.0 | 0.0 | 1.0 |
| D16S304 | 0.7 | 0.0 | 0.4 | 1.9 |
| D16S359 | 0.0 | 5.5 | 2.7 | $10^{8}$ |
| D16S416 | 1.5 | 2.8 | 2.2 | $10^{7}$ |
| D16S415 | 0.0 | 0.0 | 0.0 | 1.0 |
| D16S419 | 0.0 | 7.9 | 3.5 | $10^{18}$ |
| D16S390 | 0.5 | 4.9 | 2.7 | $10^{10}$ |
| D16S408 | 0.3 | 0.0 | 0.0 | 1.0 |
| D16S531 | 0.4 | 7.2 | 3.7 | $10^{18}$ |
| D16S320 | 0.0 | 4.8 | 2.2 | $10^{17}$ |
| D16S310 | 0.0 | 1.9 | 1.0 | $10^{3}$ |
| D16S389 | 0.0 | 2.4 | 1.5 | $10^{4}$ |
| D16S451 | 2.4 | 0.3 | 1.3 | $10^{3}$ |
| D16S265 | 0.0 | 0.0 | 0.0 | 1.0 |
| D16S164 | 2.4 | 2.1 | 2.1 | $10^{14}$ |
| D16S186 | 0.0 | 0.0 | 0.0 | 1.0 |
| D16S347 | 0.0 | 1.0 | 0.6 | $10^{5}$ |
| D16S421 | 0.0 | 0.0 | 0.0 | 1.0 |
| D16S398 | 0.0 | 0.3 | 0.0 | 1.0 |
| D16S318 | 0.0 | 0.3 | 0.0 | 1.0 |
| D16S397 | 0.0 | 0.3 | 0.6 | 1.9 |
| D16S301 | 1.1 | 0.4 | 0.7 | $10^{5}$ |

**Table 4-6.** Continued (2)

| Locus Interval | Male | Female | Sex-average | Odds(1:) |
|---|---|---|---|---|
| D16S522 | 4.3 | 1.3 | 3.2 | $10^{15}$ |
| D16S450 | 5.1 | 6.9 | 5.3 | $10^{15}$ |
| D16S395 | 0.0 | 3.0 | 2.1 | $10^{2}$ |
| D16S266 | 3.5 | 10.6 | 7.1 | $10^{11}$ |
| D16S289 | 7.6 | 6.6 | 7.1 | $10^{19}$ |
| D16S422 | 0.0 | 0.0 | 0.0 | 1.0 |
| D16S363 | 1.7 | 2.2 | 1.9 | $10^{7}$ |
| D16S402 | 5.1 | 6.8 | 5.9 | $10^{18}$ |
| D16S392 | 0.0 | 4.2 | 2.0 | $10^{2}$ |
| D16S332 | 0.0 | 0.0 | 0.0 | 1.0 |
| D16S393 | 3.8 | 5.2 | 5.3 | $10^{16}$ |
| D16S449 | 9.8 | 1.1 | 4.5 | $10^{11}$ |
| D16S305 | 0.0 | 3.7 | 1.1 | $10^{2}$ |
| D16S413 | 10.8 | 0.4 | 6.3 | $10^{11}$ |
| D16S44 | 0.0 | 0.0 | 0.0 | 1.0 |
| D16S303 | | | | |

4.4.1. The Physical Map of Chromosome 16 Based on STRs

The physical map of chromosome 16 consists of **85** STR markers. **9** (AC)n repeat markers (6 of these were used for construction of the genetic map) isolated from the random approach and **23** (AC)n repeat markers (21 of them were used for construction of the genetic map) isolated from the specific approach were physically mapped to 23 hybrid breakpoint intervals (Figure 4-3). In addition, **20** STR markers isolated by colleagues and **33** STR markers from other laboratories were physically mapped to 26 breakpoint intervals using the hybrid panel. These 85 markers were mapped into 37 breakpoint intervals of a high-resolution cytogenetic map (Figure 4-3). This cytogenetic map is one of the highest resolution physical maps available for any of the autosomes (Callen et al., 1992; Shen et al., submitted).

The region around *FRA16A* contains duplicated DNA sequences *D16S79A* and *D16S79B* (Callen et al. 1992). *D16S79A* was located between *FRA16A* and the hybrid breakpoint CY185, and *D16S79B* was located between the breakpoint intervals CY11 and CY163 (Figure 4-3). 16AC7.22 detected two loci which were designated as 16AC7.22A and B. These two loci were located in the same duplicated regions as *D16S79A* and *16S79B* (Figure 4-3). 16AC1E5, which detected two loci on chromosome 16, was located in the same region as *D16S79B* and 16AC7.22B (Figure 4-3). The (AC)n repeat *D16S312,* which also detected two

and 16AC7.22B (Figure 4-3). The (AC)n repeat *D16S312,* which also detected two loci, was located one breakpoint interval above *D16S79A* and 16AC7.22A (Figure 4-3). The localizations of these (AC)n repeats indicated that the duplicated sequence region around *FRA16A* was between hybrid breakpoints CY11 and CY19 (Figure 4-3).

The randomly isolated marker 16AC7.59 and the specifically isolated marker 16AC39C8 detected multiple loci on chromosome 16 as well as a few other chromosomes ( Figure 4-1 and 4-2). The primer sequences of these markers were part of *Alu* sequences. The GenBank data shows the primer sequences have 70-100% homology to the intron sequences of more than 10 human genes and cDNAs. *Alu* sequences have been implicated as sites of recombination events responsible for the duplication and evolution of new genes in the human genome (Zuliani and Hobbs, 1990). The multiple loci of the 16AC7.59 and 16AC39C8 seen on human chromosome 16 and chromosome 12, 14, 18 and Y could be caused by the (AC)n repeat being adjacent to *Alu* sequences, which led to duplications.

### 4.4.2. Genotyping of STR Markers

In the genotyping of 36 STR markers, a total of 12,728 genotypes were obtained from genotyping 530 individuals in 40 CEPH reference families. There are many advantages of using STR markers for genotyping, and this has been obviously demonstrated by the large number of genotypes undertaken in this study. For this

use, STR markers are easier, faster, consume less DNA and are more informative than RFLP markers.

Of 16 STR markers genotyped on the 8 largest CEPH families, 6 markers [6/47 ( 13%), 47 STRs in the framework map] were able to be incorporated in the framework map and 10 markers [10/32 (31%), 32 STRs with odds less than 1,000:1) were in the comprehensive map (Figure 4-6). The genotyping of STR markers in these 8 large families appears in some instances to be insufficient for unique ordering of these markers by linkage analysis. The total of 38 STR markers genotyped on the 8 largest CEPH families, which combined these 16 markers and the 22 markers from version 6 of the CEPH database [contributed by Weissenbach et al. (1992)], showed the same trend of insufficiency in some instances (Shen et al., submitted).

Five mutation events detected in the genotyping of 36 STR markers, in addition to 9 mutations detected in other 19 STR markers of chromosome 16 (a total of 55 STR markers were genotyped in the candidate's laboratory), the mutation rate estimated from a total of 14 mutation events was $4.0 \times 10^{-4}$ per locus per gamete per generation (Shen et al. submitted). The mutation rates of simple tandem repeats estimated by Weber and Wong (1993) and Weissenbach et al. (1992) are $1.2 \times 10^{-3}$ and $1.0 \times 10^{-3}$, respectively. The mutation rate of $4.0 \times 10^{-4}$ observed in this study is about 2.5-3 times less than those reported previously.

Heritable unstable repeat sequences have been seen in several human disorders including the fragile X syndrome (Kremer et al. 1991), myotonic dystrophy (Brook et al. 1992), spinal and bulbar muscular atrophy (Kennedy disease, La Spada et al. 1991), Huntington disease (MacDonald et al. 1993) and Spinocerebellar ataxia type 1 (SCA1, Orr et al. 1993). The instability of simple repeat sequences have also been seen in certain forms of colon cancer cells (Aaltonen et al. 1993; Thibodeau et al. 1993). Strand et al. (1993) investigated the destabilization of tracts of simple repetitive DNA in yeast. Their results indicated that DNA polymerase *in vivo* has a very high rate of slippage on templates containing simple repeat sequences, but most of these errors are corrected by cellular mismatch repair systems. Thus, the instability of simple repeat sequences (the loss or gain of repeat copy numbers) may be a consequence of either an increased rate of DNA polymerase slippage or a decreased efficiency of mismatch repair (or both).

All of the mutations in this study were detected in DNA from transformed lymphoblastoid cell lines. Royle et al. (1993) reported abnormal allele segregation of VNTR markers on chromosome 19q and X in CEPH pedigree DNAs (pedigree 12 and 1345) which were extracted from lymphoblastoid cell lines. Two abnormal segregants involved the somatic loss of the chromosome 19q distal region and of most or all of the entire paternal X chromosome, respectively. The mechanisms responsible for these allele losses probably include mitotic nondisjunction, giving rise to cells hemizygous for one of the X chromosomes, and either an interstitial or

terminal deletion of one copy of the chromosome 19, or a mitotic recombination event on the chromosome 19 (Royle et al., 1993). In this study, it is more likely that the abnormal segregations of *D16S313* and *D16S301* in two individuals were due to a mitotic mutation event or deletion event in a particular region of chromosome 16, since no abnormal allele segregations of other markers of chromosome 16 were found in these two individuals. The mutations at *D16S289* (16AC7.46), *D16S393* (16AC323H4) and *D16S310* (MIT-MH20) appear to involve the loss or gain of a small copy number of repeats. The STR mutation events of CEPH pedigree DNAs from lymphoblastoid cell lines were observed by Weber and Wong (1993) in 31 mutation events which were retested using untransformed lymphocyte cells, 19 (61%) were *in vitro* events. These reports obviously demonstrated that the somatic deletion and mutation occurred during or after the cell line was established.

Possible genotyping errors were detected as double recombinants using the computer program CHROMPIC. Nine double recombinants which could not be excluded by regenotyping were retained in the linkage data, in addition to 4 other double recombinants (which were detected from 4 STR markers isolated by other laboratories and genotyped in the candidate's laboratory) (Table 4-4), a total of 13 double recombinants detected from genotyping 55 STR markers (Shen et al. submitted). If the 13 double recombinants were in fact new mutations and combined with the 14 mutations, this would lead to a mutation rate of $7.7 \times 10^{-4}$ per locus per gamete per generation. The mutation rate was observed not to be constant from one

marker to another, and published estimations vary from $10^{-3}$-$10^{-5}$ (Weissenbach, 1993). The mutation rate in this study is within the variation range.

Null alleles were not found from the genotyping of 36 STR markers. Callen et al. (1993) reported 7 null alleles found in genotyping 7 other STR markers which were isolated in the candidate's laboratory. The null alleles were recognized by the apparent noninheritance of a parental allele by a sib. In linkage analysis or in application to diagnosis, the presence of a segregating null allele will not corrupt the linkage data (Phillips et al., 1993) but could result in loss of information.

### 4.4.3. The Genetic Linkage Map of Chromosome 16 Based on STRs

This PCR-based genetic linkage map of human chromosome 16 was constructed from 79 STR markers, 1 VNTR marker and 1 RFLP marker. These 2 non-STR markers were chosen because they extend the map towards the telomeres. This map covers the entire length of chromosome 16 , from *D16S85*, within 170-430 kb of the telomere on the short arm (Wilkie et al., 1991), to *D16S44*, within 230 kb of the telomere on the long arm (Kozman et al., 1993). The median distance and the average distance between markers on the framework map is 2.7 and 3.2 cM, respectively. In comparison with the averaged resolutions of other STR-based linkage maps of 8 cM (chromosome 1, 38 STRs, Engelstein et al., 1993), 15 cM (chromosome 4, 16 STRs, Mills et al., 1992), 5.8 cM (chromosome 11, 25 STRs, Litt et al., 1993), 9.4 cM (chromosome 12, 22 STRs, Dawson et al., 1993), 8.2 cM

(chromosome 13, 21STRs, Petrukhin et al., 1993), 7.3 cM (chromosome 18, 14 STRs, Straub et al., 1993) and 8.5 cM (chromosome 20, 26STRs, Hazan et al., 1992), 4.6 cM (chromosome 22, 24 STRs, Buetow et al. 1993), the resolution of this map is much higher [recently, McInnis et al. (1993) reported an average distance between markers on the genetic map of chromosome 21 based on 43 STRs is 2.5 cM; Beckmann et al. (1993) reported a linkage map of chromosome 15q with an average distance of 2 cM and containing 55 polymorphic STR markers.].

The framework and comprehensive maps were anchored to the high-resolution cytogenetic map (Figure 4-6). It is apparent that the combination of genetic linkage analysis and physical mapping can be extremely helpful in resolving locus order at the resolution of the comprehensive map. For example, the genetic location of marker *D16S452* is between *D16S287* and *D16S420* on the short arm, but the physical location is between hybrid breakpoints CY11 and CY180(P). Therefore the genetic location can be reduced to the interval between *D16S410* and *D16S420* (Figure 4-6). For eleven comprehensive map markers (*D16S94, D16S523, D16S405, D16S452, D16S383, D16S261, D16S411, D16S164, D16S347, D16S266* and *D16S363* ), the genetic distance of location interval was narrowed by physical mapping (Figure 4-6).

The sex-specific maps indicate the difference between the male and female recombination rates (Table 4-5). The female map is 1.4 times longer than the male

map. Between the distal regions of the chromosome arms, the greater recombination rate on the female map exhibits the same general trends as previous reports (Donis-Keller et al., 1987; Julier et al., 1990; Keith et al., 1990; Kozman et al., 1993). In contrast to the general phenomenon of a greater recombination rate in females, the recombination frequencies in two distal regions, *D16S406* to *D16S85* on the short arm and *D16S449* to *D16S303* on the long arm, show greater recombination in males (Table 4-5). The genetic linkage maps of chromosome 16 published by Donis-Keller et al. (1987), Julier et al. (1990), Keith et al. (1990) and Kozman et al. (1993) also observed the same expansion of male genetic distance in telomeric regions.

The physical length of the centromeric region (including heterochromatic region) of chromosome 16 is 6-9 Mb (Doggett, 1992). The genetic distance between the closest markers either side of the centromere (*D16S300* and *D16S285* ) is 1.0 cM on the sex-averaged map (the genetic distance on the female map is 1.9 cM. No recombination event was observed in males) (Table 4-6 and Figure 4-6). The physical distance corresponding to 1 cM is, on average, 1 Mb throughout the genome. This suggests there is a possible suppression of recombination over the centromeric region (including heterochromatic region), since its physical length is about 6-9% of the whole of chromosome 16 (The physical length of chromosome 16 is about 98 Mb, Morton, 1991).

Wevrick and Willard (1989) investigated the meiotic segregation of multiple α satellite repeats at the centromeric region of human chromosomes 1, 7, 10, 11, 16, 17, X and Y. An average of fourteen informative meioses per autosomal centromere were examined, representing the segregation of about 191,000 kb of α satellite through 84 meiotic events. No recombination events or meiotic rearrangements were observed. This result suggested that 1) α satellite repeats were meiotically stable; 2) the estimate of recombination frequency (based on 95% confidence limits) was 0-1.5% meiotic recombination per megabase of α satellite DNA. In the study of chromosome 16, no recombination event was observed in males and 2.0% recombination rate was estimated in females. The recombination events observed from the two flanking markers on chromosome 16 showed the same phenomenon of low recombination rate over the centromeric region. Wevrick and Willard indicated that further analysis was required to determine more precisely the rates of recombination and whether recombination rates were suppressed at human centromeric regions as was the case in other organisms (Kaback et al., 1989). Pardue et al. (1987) and Stallings et al. (1991) have observed a lack of (AC)n repeat sequences in the centromeric heterochromatic region of human chromosomes. In this study no (AC)n repeat sequences were isolated from the centromeric heterochromatic region of chromosome 16. The rate of meiotic recombination of (AC)n repeat markers in this region, therefore, can only be estimated by flanking markers. In order to compare genetic and physical distances in the centromeric

region of human chromosomes, and to confirm whether recombination suppression exists in the centromeric region of human chromosome 16, the investigation of more informative polymorphic DNA markers in this region would be required to detect meiotic recombinations.

The comprehensive map covers the entire chromosome 16 from *D16S85* to *D16S44*, and gives the genetic length of 126.8 cM in males, 178.9 cM in females, and an average of 151.1 cM in both sexes. By comparison, the genetic lengths of this map are in agreement with those estimated from published genetic and chiasma chromosome 16 maps (Table 4-7). The lengths of multipoint linkage maps are affected by the choice of mapping function and the frequency of errors (Morton and Collins, 1990; Buetow, 1991). Possible genotyping errors detected as double recombinants using CHROMPIC were checked against the autoradiographs and by regenotyping the original DNA samples. After this error checking, 13 double recombinants remained in the linkage data because they could not be excluded by regenotyping. The similar lengths of this genetic map to those previously reported indicate that the length of the map is apparently not affected by these double recombination events or that the error rate in this map is similar to that of previous maps.

The distribution of simple tandem repeats is non-random on both the physical map and genetic linkage maps (Figure 4-6) with a number of clusters and gaps.

**Table 4-7.** Estimates of Total Map Lengths for Human Chromosome 16

| Male (cM) | Female (cM) | Sex-Averaged (cM) | Reference |
|---|---|---|---|
| 164 | 237 | 194 | Donis-Keller et al. 1987 |
| 186 | 226 | | Julier et al. 1990 |
| 115 | 193 | 149 | Keith et al. 1990 |
| 120 | 193 | 157 | Morton et al. 1991 |
| 111 | 179 | 145 | Morton et al. 1991 (Chiasma data) |
| 133 | 197 | 162 | NIH/CEPH 1992 |
| | | 116* | Weissenbach et al. 1992 |
| 133 | 202 | 165 | Kozman et al. 1993 |
| 127 | 179 | 151 | This Study |

* This map did not extend the full length of human chromosome 16.

Nine breakpoint intervals on the physical map contain three or more STR markers. The average interlocus distance in seven clusters ranged from 0.10 to 0.88 cM. One interval on the short arm flanked by the CY165 and CY15 hybrid breakpoints contains 9 STR markers with an average interlocus distance of 1.7 cM. Another interval on the distal long arm flanked by CY100, CY120 and CY18A(P) breakpoints contains 6 STR markers with average interlocus distance of 3.3 cM. The largest gap of 13.6 cM, between *D16S405* and *D16S414*, is the only gap in excess of 10 cM. This gap is located at the interval flanked by CY19 and CY180(D) breakpoints on the physical map. This interval appears to be deficient in simple tandem repeats. Weissenbach et al. (1992) also observed non-uniform distribution of (AC)n repeats using random isolation from the whole human genome. There are several explanations of this non-random distribution (Weissenbach et al., 1992; Weissenbach, 1993): 1) expansion of genetic distance of subtelomeric regions; 2) difficulties in PCR amplification of GC rich regions; 3) source of libraries for isolation of simple tandem repeats constructed using size fractionated fragments which were digested by specific restriction enzymes; 4) exclusion of less informative STR markers and 5 ) primers could not be made for some STR markers because of the limited flanking sequences. All of these could cause the observed clusters and gaps on the physical and genetic maps of chromosome 16. When more cosmids and YACs are physically mapped on chromosome 16, the STR markers could be isolated by a specific approach to fill the gaps or the regions of interest.

# CHAPTER 5

## LOCALIZATION OF THE *MEF* GENE

## 6.1. INTRODUCTION

### 6.1.1. Familial Mediterranean Fever (FMF)

Familial Mediterranean fever (FMF), also known as recurrent polyserositis, is an autosomal recessive disorder of unknown aetiology. According to McKusick (1992, No.249100), many cases were described from Lebanon (most of them were Armenian) in 1954. It is characterized by recurrent, self-limited attacks of fever, with peritonitis, arthritis and pleuritis, accompanied by pain in the abdomen, chest or joints and an erysipelas-like erythema. Amyloidosis is the most severe complication of the disease, which led to death from renal failure before the introduction of colchicine prophylaxis (McKusick, 1992). This disease is relatively common in individuals of Mediterranean descent, especially in non-Ashkenazi Jews, Armenians, Anatolian Turks and Middle Eastern Arabs. The frequency of the disease gene among these populations is extraordinarily high, reaching 1 in 22 among Jews in North Africa and 1 in 14 among Armenians in Los Angeles (Rogers et al., 1989). Although the biochemical basis of this disease is unknown, the clinical manifestations of FMF suggest a lesion in a molecule important to the understanding of inflammation in general (Aksentijevich et al., 1993b).

### 6.1.2. Background of Localization of the *MEF* Gene

During the past several years, there has been an ongoing effort to map the gene responsible for FMF [designated by the gene symbol "*MEF*" (Aksentijevich et al., 1993b)] by molecular genetic techniques. The serum amyloid A gene was initially suspected to be abnormal in FMF patients (Sack et al., 1988). To test the role of serum amyloid A (SAA) and P (APCS) in FMF patients, Shohat et al. (1990a) studied 17 informative families (15 Armenians and 2 non-Ashkenazi Jews) and 8 FMF patients with amyloidosis using a candidate gene approach. Their data showed no evidence for any FMF-associated polymorphism in any of 41 Armenian and Jewish FMF patients. Their results excluded close linkage between SAA and *MEF* (lod score = -2.16), and between APCS and *MEF* (lod score = -2.2). Shohat et al. indicated that SAA and APCS were not the genes causing FMF and defects in them did not seem to contribute to the pathogenesis of FMF-amyloidosis. Sack et al. (1991) also studied the relationship between FMF and the SAA gene (located on chromosome 11p) by typing alleles of an (AC)n repeat and a RFLP in the SAA gene cluster in Israeli FMF kindreds. The result from linkage analysis eliminated a minimum of 10.4 cM including and surrounding the SAA gene cluster as the site of the FMF mutation.

Shohat et al. (1990b) screened 14 Armenian families for 19 polymorphic markers (4 red cell antigens, 10 red cell enzymes and 5 serum proteins). These

markers were located on 12 human chromosomes. The results showed no association found with any markers studied. By linkage analysis, 14 markers were excluded and 5 markers (one of them was phosphoglycolate phosphatase located on chromosome 16p13) could not be excluded.

Aksentijevich et al. (1991) performed linkage analysis on 18 FMF families, which were collected in Isreal, most of them were of North African and Iraqi descent. Linkage analysis with 95 DNA markers excluded about 30% of the human genome, with most of chromosomes 1, 9, 10 and 15 being ruled out. Four markers (*D17S74*, *D17S40*, *D17S35* and *GH*) on the long arm of chromosome 17 appeared to be linked to the FMF susceptibility gene. Multipoint linkage analysis yielded a lod score of 3.54 at a 15 cM map distance from *D17S40*. In this report Aksentijevich et al. presented a most likely map order: *D17S74 - GH - D17S35/D17S40 - MEF*.

After finding linkage between the markers on chromosome 17q and the *MEF* gene, Kastner's group (Pras et al., 1992; ) and Shohat's group (Shohat et al., 1992) in USA independently reported that the *MEF* gene is linked to the markers on chromosome 16p13.3. Pras et al. (1992) found the marker *D16S84* gave a maximal lod score of 9.17 at $\theta = 0.04$, and the marker *D16S85* (5'HVR), which is associated with the haemoglobin $\alpha$-chain complex, gave a maximal lod score of 14.47 at $\theta = 0.06$. The *MEF* gene was mapped proximal to these two markers.

Because there were clinical differences between the Armenian FMF patients and the non-Ashkenazi Jewish patients, the possibility of genetic locus heterogeneity needed to be examined. For this purpose, Shohat et al. (1992) tested FMF families from both Armenian and non-Ashkenazi Jewish populations with the markers on chromosomes 16p and 17q. Their results indicated that the gene for FMF was linked to the α-globin complex on chromosome 16p13.3 in both Armenians and non-Ashkenazi Jews. This gave evidence for locus homogeneity. In the report of Aksentijevich et al. (1993a), the previous localization of *MEF* on chromosome 17q (Aksentijevich et al., 1991) was proved to be a "false positive" (type I) error. They indicated that chromosome 17q did not encode a major FMF susceptibility gene for some of the families, nor did it encode a disease-modifying gene (second locus).

Since the *MEF* gene was known to be linked to the markers on chromosome 16p13 in 1992, and (AC)n repeat markers were isolated in that region, further refining of this gene localization became feasible. Recently Aksentijevich et al. (1993b) and Fischel-Ghodsian et al. (1993) used four (AC)n repeat markers and six RFLP markers on 16p13 to refine the mapping of *MEF*. Aksentijevich et al. placed the *MEF* gene between the RFLP marker 24.1 (*D16S80*) on the centromeric side and the (AC)n marker VK5 (*D16S94*) on the telomeric side, in a genetic interval of about 9 cM. Fischel-Ghodsian et al. placed the *MEF* gene between the RFLP marker 24.1 (*D16S80*) and the (AC)n marker SM7 (*D16S283*), at a distance of approximately 3 cM centromeric from SM7.

### 6.1.3. Aim of the Project

Localization of *MEF* gene was a project based on collaboration with Dr. D.L. Kastner's group in National Institute of Arthritis and Musculoskeletal and Skin Diseases (Bethesda, USA). It started in the early of 1992, when Kastner's group found the linkage between the markers [*D16S84* (RFLP) and *D16S85* (VNTR)] and the *MEF* gene on chromosome 16p13.3 (Pras et al., 1992). The aim of the project was to refine localization of *MEF* gene using (AC)n repeat markers by analysis of recombinant haplotypes, homozygosity mapping and linkage analysis. The (AC)n repeat markers were selected from the genetic linkage map of chromosome 16, which was constructed using simple tandem repeat polymorphisms (chapter 4).

## 6.2. MATERIALS AND METHODS

### 6.2.1. DNA Samples and FMF Pedigrees

A total of **372** DNA samples from **62** FMF pedigrees were provided by Dr. D. L. Kastner and Dr. N. Fischel-Ghodsian (Department of Paediatrics of Cedars-Sinai Medical Centre, Los Angeles, USA ) (see appendix II). Forty-nine FMF families with **303** individuals (175 affected) from Dr. Kastner were non-Ashkenazi Jews. They were Moroccan, Libyan, Tunisian, Egyptian, Iraqi, Kurdish or Greek origins. Of 49 families, 18 involved consanguineous marriages. Thirteen families with **69** individuals (28 affected) from Dr. Fischel-Ghodsian were Armenian

(Shohat, 1992). DNA was extracted by standard techniques. The diagnosis of FMF was made according to established clinical criteria (Sohar et al., 1967).

6.2.2. Simple Tandem Repeat Markers

Four (AC)n repeat markers (*D16S291, D16S94, D16S523* and *D16S453*) were used for this study. *D16S291* (AC2.5) (Thompson et al., 1992) and *D16S94* (ACVK5) (Aksentijevich et al., 1993b) were isolated by other members of the candidate's laboratory. *D16S523* (AC13H1) (Shen et al., 1993d) and *D16S453* (AC301G12) (Shen et al., 1993c) were isolated by the candidate. Characterization of these four markers were shown in Table 6-1. Heterozygosities ranged from 0.45 to 0.79. Four markers were physically mapped at 16p13.3 using a hybrid panel (Figure 4-3). *D16S291* and *D16S94* were located in one breakpoint interval (between CY14, CY192 and CY190, CY186, 23HA). *D16S523* and *D16S453* were located in one interval below (between CY190, CY186, 23HA and CY177, CY194, CY182). These four markers were genetically mapped in a 4.9 cM region of 16p13.3 on the CEPH background map (Table 4-6). The order of marker loci is: 16pter -- *D16S291* $\underline{1.1cM}$ *D16S283* $\underline{0.4cM}$ *D16S94* $\underline{1.6cM}$ *D16S523* $\underline{1.8cM}$ *D16S453* --- 16cen (Table 4-6, *D16S283* data was contributed by Kastner's group).

**Table 6-1.** Characterization of Four (AC)n Repeat Markers (CEPH Data)

| Locus | Clone | Repeat | No. Alleles(bp) | Het | Reference |
|---|---|---|---|---|---|
| D16S291 | AC2.5 | (GT)25 | 9(154-170) | 0.79 | Thompson et al. 1992 |
| D16S94 | ACVK5B | (GT)20 | 5(82-90) | 0.51 | Aksentijevich et al. 1993b |
| D16S523 | AC13H1 | (GT)19+(GT)3 | 5(77-87) | 0.68 | Shen et al. 1993d |
| D16S453 | AC301G12 | (GT)11+(GT)4 | 2(125-127) | 0.45 | Shen et al. 1993c |

6.2.3. Genotyping of STR markers

Initially *D16S291* (AC2.5) and *D16S94* (ACVK5) were genotyped through the first set of **192** DNA samples of **31** FMF families from Kastner (Table 6-2 and Appendix II). *D16S523* (AC13H1) and *D16S453*(AC301G12) were genotyped through **372** DNA samples of **62** FMF families (Table 6-2 and Appendix II). These (AC)n markers were genotyped using PCR conditions as described in chapter 4.

6.2.4. Statistical Analysis for Detection of Allele Association

Linkage disequilibrium studies have been used to support the mapping of disease genes and for the detection of founder effects (Richards et al., 1992; Mitchison et al., 1993; Aksentijevich et al., 1993c). The association of FMF susceptibility alleles at *D16S523* and *D16S453* were analysed in 14 Moroccan Jewish FMF families, 16 non-Moroccan Jewish FMF families and 13 Armenian FMF families (Table 6-4, 6-5, 6-6) [The allele association of susceptibility alleles at *D16S291* and *D16S94* has been published (Aksentijevich et al., 1993c)]. For each pedigree, the parental alleles associated with FMF chromosomes and non-FMF chromosomes were identified. $\chi^2$ tests were used to determine the significance of the allele distributions observed. For evaluation of linkage disequilibrium, the Yule association coefficient (Yule and Kendall, 1968) was calculated according to the formula

**Table 6-2.** Genotyping of (AC)n Repeat Markers on 62 FMF Families

| FMF Families[a] | D16S291 (AC2.5) | D16S94 (ACVK5) | D16S523 (AC13H1) | D16S453 (AC301G12) |
|---|---|---|---|---|
| #2* | +[b] | + | + | + |
| #3** | + | + | + | + |
| #4** | + | + | + | + |
| #6* | + | + | + | + |
| #7* | + | + | + | + |
| #8* | + | + | + | + |
| #10* | + | + | + | + |
| #11** | + | + | + | + |
| #12* | -- | -- | + | + |
| #13* | + | + | + | + |
| #14* | + | + | + | + |
| #15* | + | + | + | + |
| #16** | + | + | + | + |
| #17** | + | + | + | + |
| #21** | + | + | + | + |
| #23* | + | + | + | + |
| #25* | + | + | + | + |
| #27** | + | + | + | + |
| #29* | -- | -- | + | + |
| #30** | + | + | + | + |
| #31* | + | + | + | + |
| #33* | + | + | + | + |
| #35** | + | + | + | + |
| #38** | + | + | + | + |
| #39** | + | + | + | + |
| #40/62** | + | + | + | + |
| #42* | + | + | + | + |
| #46* | + | + | + | + |
| #49** | + | + | + | + |
| #60* | + | + | + | + |
| #72* | + | + | + | + |
| #75** | + | + | + | + |
| #101 | -- | -- | + | + |
| #102 | -- | -- | + | + |
| #104 | -- | -- | + | + |
| #105 | -- | -- | + | + |
| #106 | -- | -- | + | + |

a #2-145 (a total of 49 FMF fmilies) from Dr. D. L. Kastner were non-Ashkenazi. F1-14 (a total of 13 FMF families) from Dr. N. Fichel-Ghodsian were Armenian.
b "+" was typed and "--" was untyped.
* 16 non-Moroccan FMF families.
** 14 Moroccan FMF families.

**Table 6-2.** continued

| FMF Families[a] | D16S291 (AC2.5) | D16S94 (ACVK5) | D16S523 (AC13H1) | D16S453 (AC301G12) |
|---|---|---|---|---|
| #108 | -- | -- | + | + |
| #109 | -- | -- | + | + |
| #112 | -- | -- | + | + |
| #115 | -- | -- | + | + |
| #117 | -- | -- | + | + |
| #118 | -- | -- | + | + |
| #119 | -- | -- | + | + |
| #121 | -- | -- | + | + |
| #123 | -- | -- | + | + |
| #125 | -- | -- | + | + |
| #130 | -- | -- | + | + |
| #145 | -- | -- | + | + |
| F1 | -- | -- | + | + |
| F2 | -- | -- | + | + |
| F3 | -- | -- | + | + |
| F4 | -- | -- | + | + |
| F5 | -- | -- | + | + |
| F5 | -- | -- | + | + |
| F6 | -- | -- | + | + |
| F7 | -- | -- | + | + |
| F9 | -- | -- | + | + |
| F10 | -- | -- | + | + |
| F11 | -- | -- | + | + |
| F12 | -- | -- | + | + |
| F13 | -- | -- | + | + |
| F14 | -- | -- | + | + |

$$| A | = | (ad - cb) / (ad + cb) |,$$

where a is the number of non-FMF chromosomes with allele A, b is the number of FMF chromosomes with allele A, c is the number of non-FMF chromosomes with allele B, and d is the number of FMF chromosomes with allele B.

### 6.2.5. Linkage Analysis

Linkage analysis was performed using LINKAGE analysis package (Version 5.1) (Lathrop and Lalouel, 1984). LINKAGE is written in Turbo-Pascal Version 5 (Borland International Inc.) and run on an IBM compatible personal computer.

LINKAGE used the files INFILE, PEDFILE and DATAFILE for input of data. INFILE contains pedigree structure, phenotype and genotype information for disease and marker loci. PEDFILE is a modified form of INFILE which includes offspring and its parents. DATAFILE contains information about disease and marker loci, such as allele frequencies and penetrance.

The LINKAGE package of programs analyses general pedigrees of arbitrary structure for linkage and genetic risk.

MAKEPED is the program which converts INFILE to PEDFILE.

PREPLINK is the program which constructs the DATAFILE.

LINKAGE CONTROL PROGRAM (LCP) is the program which builds command files to perform linkage analysis using MLINK.

MLINK is the program which calculates pairwise lod score for a given recombination fraction between two loci and for risk analysis using the PEDFILE and DATAFILE files.

6. LINKAGE REPORT PROGRAM (LRP) is the program which displays and formats the stream out put from the linkage analysis program.

In this study, MLINK was used for two-point linkage analysis. INFILE and DATAFILE were generated using FMF genotype data from four (AC)n repeat markers (*D16S291, D16S94, D16S523, D16S453*) and used for entry of data to MLINK. Pairwise lod scores between the disease locus and marker loci were calculated at recombination fractions of 0.0, 0.01, 0.05, 0.10, 0.20, 0.30, and 0.40 using MLINK. The maximun lod scores at sex-averaged recombination fractions were also calculated using MLINK. The penetrance used was 95% in males, 70% in females (Aksentijevich et al., 1993b).

Allele frequencies (Table 6-3) for four (AC)n marker loci were determined by genotypes of unrelated individuals of FMF families. Allele frequencies for the disease locus were 0.9550 for the normal allele and 0.0450 for the affected allele (Aksentijevich et al., 1993b).

Twelve of 18 consanguineous families [the exact relationship of the partners in 6 families (#27, #38, #104, #109, #115, #119) (Appendix II) were unknown] were broken inbred loops at the father's side using program MAKEPED.

The two-point linkage analysis between the *MEF* locus and each of the three (AC)n markers [*D16S94* (ACVK5), *D16S523* (AC13H1) and *D16S453* (AC301G12)] was performed by the candidate. *D16S523* and *D16S453* were analysed in 62 FMF families (non-Ashkenazi Jews and Armenian), and *D16S94* was analysed in 31 FMF families (non-Ashkenazi Jews). The two-point analysis of *D16S291* (AC2.5) was performed in 31 FMF families (non-Ashkenazi) in Kastner's laboratory (Aksentijevich et al., 1993b). Multipoint linkage analysis for the *MEF* gene locus plus the markers *D16S246* (218EP6) and *D16S523* (13H1) was analysed in 49 non-consanguineous FMF families and 11 consanguineous FMF families, and was performed in Kastner's laboratory. The sex-average recombination fraction between *D16S246* and *D16S523* was taken to be $\theta = 0.02$ whcih was obtained from two-point linkage analysis.

## 6.3. RESULTS

### 6.3.1. Genotyping Data

Three hundred and seventy two genotypes for each (AC)n marker were shown in Appendix II. From genotyping unrelated individuals of FMF families, the expected

allele frequencies for four markers were calculated (Table 6-3). The allele frequencies (Table 6-3) between CEPH and FMF families were similar. Two differences were 1) the 156bp allele frequency (0.08) of *D16S291* in FMF was much lower than CEPH (0.26); 2) several alleles not detected in CEPH, were seen in FMF families (Table 6-3). The alleles only seen in FMF families did not show any association with FMF ( Table 6-4, 6-5 and 6-6). These represent differences in allele frequency between the different ethnic populations.

### 6.3.2. Allelic Association

Allelic association (linkage disequilibrium) was analysed between FMF and alleles at two (AC)n repeat loci (*D16S523* and *D16S453*). The distribution of individual alleles at these two loci in the three ethnic groups (Moroccan, non-Moroccan and Armenian) shown in Tables 6-4, 6-5 and 6-6. The most common alleles of FMF carrier chromosomes in the three ethnic groups were 83bp at *D16S523* and 125bp at *D16S453*. There is no significant allelic association for these two loci among non-Moroccans and Armenians ($P > 0.05$, Table 6-5 and 6-6). In the Moroccan group the 83bp allele at *D16S523* showed significant association with FMF ($P < 0.05$ for both combined allele and multiallele calculations), but the 125bp allele at *D16S453* did not show a significant association ($P > 0.05$, Table 6-4). These two alleles were also observed to be the most common alleles in non-FMF carrier chromosomes (Tables 6-4 to 6-6) and the normal

**Table 6-3.** Allele Frequency and Heterozygosity of Four (AC)n Repeat Markers on CEPH and FMF Families

| | CEPH (160 chr.) | | | non-Ashkenazi(102 chr.) | | | Armenian (48 chr) | | |
|---|---|---|---|---|---|---|---|---|---|
| Locus | Alleles(bp) | Frequency | Het | Alleles(bp) | Frequency | Het | Alleles(bp) | Frequency | Het |
| D16S291 | | | 0.79 | | | 0.76 | | | |
| AC2.5 | 170 | 0.01 | | | | | | | |
| | 168 | 0.01 | | 168 | 0.01 | | | | |
| | 166 | 0.08 | | 166 | 0.11 | | | | |
| | 164 | 0.23 | | 164 | 0.28 | | | | |
| | 162 | 0.25 | | 162 | 0.36 | | | | |
| | 160 | 0.12 | | 160 | 0.09 | | | | |
| | 158 | 0.03 | | 158 | 0.08 | | | | |
| | 156 | 0.26 | | 156 | 0.08 | | | | |
| | 154 | 0.01 | | | | | | | |
| | | | | 152 | 0.01 | | | | |
| D16S94 | | | 0.51 | | | 0.54 | | | |
| ACVK5 | | | | 92 | 0.01 | | | | |
| | 90 | 0.02 | | 90 | 0.05 | | | | |
| | 88 | 0.35 | | 88 | 0.23 | | | | |
| | 86 | 0.62 | | 86 | 0.64 | | | | |
| | 84 | 0.02 | | 84 | 0.02 | | | | |
| | 82 | 0.01 | | 82 | 0.06 | | | | |

**Table 6-3.** Continued

| | CEPH (160 chr.) | | | non-Ashkenazi(102 chr.) | | | Armenian (48 chr) | | |
|---|---|---|---|---|---|---|---|---|---|
| Locus | Alleles(bp) | Frequency | Het | Alleles(bp) | Frequency | Het | Alleles(bp) | Frequency | Het |
| D16S523 | | | 0.68 | | | 0.62 | | | 0.61 |
| AC13H1 | 87 | 0.02 | | 87 | 0.01 | | | | |
| | 85 | 0.21 | | 85 | 0.16 | | 85 | 0.19 | |
| | 83 | 0.44 | | 83 | 0.57 | | 83 | 0.58 | |
| | | | | 81 | 0.01 | | 81 | 0.02 | |
| | 79 | 0.04 | | 79 | 0.08 | | 79 | 0.06 | |
| | 77 | 0.29 | | 77 | 0.17 | | 77 | 0.13 | |
| | | | | 75 | 0.01 | | | | |
| | | | | | | | 71 | 0.02 | |
| D16S453 | | | 0.45 | | | 0.39 | | | 0.44 |
| AC301G12 | | | | 129 | 0.04 | | 129 | 0.02 | |
| | 127 | 0.34 | | 127 | 0.21 | | 127 | 0.29 | |
| | 125 | 0.66 | | 125 | 0.75 | | 125 | 0.69 | |

**Table 6-4.** Allelic Association for 14 Moroccan Jewish FMF families

| Marker Alleles | No.(%) of FMF Chr. | | No.(%) of non-FMF Chr. | | $\chi^{2a}$ | | \| A \|[b] |
|---|---|---|---|---|---|---|---|
| | | | | | Combined Allele | Multiallele | |
| D16S523 (AC13H1) | | | | | | | |
| 87bp | 0 | (0) | 0 | (0) | | | |
| 85bp | 1 | (3.7%) | 5 | (19%) | | | |
| 83bp[c] | 21 | (78%) | 14 | (52%) | 3.98 | 10.37 | |
| 81bp | 0 | (0) | 0 | (0) | (1 df) | (3 df) | 0.53 |
| 79bp | 4 | (15%) | 1 | (3.7%) | (P< 0.05) | (P< 0.05) | |
| 77bp | 1 | (3.7%) | 7 | (26%) | | | |
| D16S453 (AC301G12) | | | | | | | |
| 129bp | 0 | (0) | 1 | (3.7%) | 2.65 | 4.10 | |
| 127bp | 4 | (14%) | 9 | (33%) | (1 df) | (2 df) | 0.56 |
| 125bp[c] | 24 | (86%) | 17 | (63%) | (P> 0.05) | (P> 0.05) | |

a $\chi^2$ test of the null hypothesis of no linkage disequilibrium, calculated as described in Methods and Materials. Comparison in each case to combine all other alleles together as one group or to consider each allele separately.
b Yule association coefficient.
c FMF-associated allele.

**Table 6-5.** Allelic Association for 16 non-Moroccan Jewish FMF families

| Marker Alleles | No.(%) of FMF Chr. | | No.(%) of non-FMF Chr. | | $\chi^{2a}$ Combined Allele | $\chi^{2a}$ Multiallele | \| A \|[b] |
|---|---|---|---|---|---|---|---|
| D16S523 (AC13H1) | | | | | | | |
| 87bp | 0 | (0) | 1 | (3.7%) | | | |
| 85bp | 5 | (19%) | 3 | (11%) | 0.08 | 2.94 | |
| 83bp[c] | 15 | (56%) | 16 | (59%) | (1 df) | (4 df) | 0.12 |
| 81bp | 0 | (0) | 0 | (0) | (P> 0.05) | (P> 0.05) | |
| 79bp | 1 | (3.7%) | 3 | (11%) | | | |
| 77bp | 6 | (22%) | 4 | (15%) | | | |
| D16S453 (AC301G12) | | | | | | | |
| 129bp | 4 | (14%) | 1 | (3.7%) | 0.34 | 1.95 | |
| 127bp | 3 | (11%) | 4 | (15%) | (1 df) | (2 df) | 0.19 |
| 125bp[c] | 21 | (75%) | 22 | (81%) | (P> 0.05) | (P> 0.05) | |

a $\chi^2$ test of the null hypothesis of no linkage disequilibrium, calculated as described in Methods and Materials. Comparison in each case to combine all other alleles together as one group or to consider each allele separately.
b Yule association coefficient.
c FMF-associated allele.

**Table 6-6.** Allelic Association for 13 Armenian FMF families

| Marker Alleles | No.(%) of FMF Chr. | | No.(%) of non-FMF Chr. | | $\chi^{2a}$ Combined Allele | $\chi^{2a}$ Multiallele (p) | \| A \|[b] |
|---|---|---|---|---|---|---|---|
| D16S523 (AC13H1) | | | | | | | |
| 87bp | 0 | (0) | 0 | (0) | | | |
| 85bp | 8 | (32%) | 2 | (8.7%) | | | |
| 83bp[c] | 12 | (48%) | 16 | (70%) | 0.89 | 4.54 | 0.42 |
| 81bp | 1 | (4.0%) | 0 | (0) | (1 df) | (3 df) | |
| 79bp | 1 | (4.0%) | 2 | (8.7%) | (P> 0.05) | (P> 0.05) | |
| 77bp | 2 | (8.0%) | 4 | (13%) | | | |
| 75bp | 0 | (0) | 0 | (0) | | | |
| 73bp | 0 | (0) | 0 | (0) | | | |
| 71bp | 1 | (4.0%) | 0 | (0) | | | |
| D16S453 (AC301G12) | | | | | | | |
| 129bp | 1 | (4%) | 0 | (0) | 2.12 | 3.91 | 0.44 |
| 127bp | 4 | (16%) | 9 | (39%) | (1 df) | (2 df) | |
| 125bp[c] | 20 | (80%) | 14 | (61%) | (P> 0.05) | (P> 0.05) | |

a $\chi^2$ test of the null hypothesis of no linkage disequilibrium, calculated as described in Methods and Materials. Comparison in each case to combine all other alleles together as one group or to consider each allele separately.
b Yule association coefficient.
c FMF-associated allele.

populations (CEPH families, Table 6-3). From Yule association coefficient calculation, there was no statistically significant difference in the strength of the allelic association at these loci (Table 6-4 to 6-6).

6.3.3. Recombinants

Table 6-7 is a summary of observed recombinants between the *MEF* gene and chromosome 16p13.3 markers. To establish the cosegregating region for FMF, haplotypes were constructed by the genotypes from four (AC)n repeat markers and another five markers (from Kastner and Fischel-Ghodsian, Table 6-7). The order of these markers shown in Table 6-7 is based on previous linkage analysis in FMF families (Aksentijevich et al., 1993b), the CEPH background map (Table 4-6) and personal communications (Dr. Kastner, 1993-1994).

In 62 FMF families, 35 recombinants were observed (Table 6-7). The recombinations were observed between the *MEF* locus and nine 16p13.3 markers. In family #F3, the normal individual F3-16 had a crossover between *D16S523* (AC13H1) and *D16S246* (218EP6) which places the *MEF* gene between these two markers (both markers were informative). Four other crossovers in recombinants 13-10, 112-13, 121-06 and 145-11 support the centromeric boundary for the location of *MEF* at AC13H1. Two crossovers in recombinants 72-05 and F14-59 support the telomeric boundary at 218EP6. It is seen that 218EP6 and AC13H1 are

**Table 6-7.** Observed Recombinants between *MEF* and Chromosome 16p13.3 Marker Loci

| | | Markers[a] (Centromere---Telomere) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Individual | Status** | D16S423* AFM249 | D16S63* AC327 | D16S80* AC24-1 | D16453 AC301G12 | D16S523 AC13H1 | D16S246* 218EP6 | D16S94 ACVK5 | D16S283* SM7 | D16S291 AC2.5 |
| #2-07 (maternal) | A | -- | ? | ? | ? | ? | ? | + | ? | ? |
| #2-07 (paternal) | A | + | + | ? | + | + | ? | -- | ? | + |
| #7-06 | N | -- | ? | ? | ? | -- | ? | + | + | + |
| #10-06 | A | -- | + | ? | + | + | ? | ? | + | + |
| #11-03 | A | -- | + | ? | + | + | ? | ? | + | + |
| #12-03 | A | + | -- | ? | -- | + | ? | + | + | ? |
| #13-03 (mother affected) | A | + | + | + | + | + | + | + | + | + |
| #13-10 | A | -- | -- | -- | -- | -- | ? | ? | + | + |
| #15-03 | A | + | ? | ? | + | + | + | -- | -- | -- |
| #21-05 | A | -- | ? | ? | -- | ? | + | + | + | + |
| #25-06 | A | -- | ? | -- | -- | ? | + | ? | + | + |
| #25-07 | N | -- | ? | -- | + | ? | + | + | + | + |

a + sign indicates marker derived from the affected chromosome; -- sign indicates marker derived from the unaffected chromosome; and ? indicates marker not informative or not done. * data from Dr. Kastner and Dr. Fischel-Ghodsian. ** A = affected; N = normal

**Table 6-7.** Continued (1)

| Individual | Status** | Markers[a] (Centromere---Telomere) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | D16S423* AFM249 | D16S63* AC327 | D16S80* 24-1 | D16453 AC301G12 | D16S523 AC13H1 | D16S246* 218EP6 | D16S94 ACVK5 | D16S283* SM7 | D16S291 AC2.5 |
| #27-12 | A | -- | -- | -- | ? | ? | + | + | + | + |
| #33-04 | A | + | + | + | ? | ? | + | ? | -- | -- |
| #35-02 | A | -- | + | ? | + | + | ? | + | + | + |
| #38-08 | A | + | + | + | + | + | ? | ? | -- | -- |
| #42-05 | A | ? | -- | + | + | + | + | + | + | + |
| #46-06 | A | -- | ? | + | ? | ? | ? | + | + | + |
| #62-06 | A | -- | -- | ? | ? | ? | + | + | + | + |
| #72-05 | A | + | + | ? | ? | ? | -- | -- | -- | -- |
| #102-05 | A | -- | -- | -- | ? | ? | ? | + | + | ? |
| #104-11 | N | + | -- | -- | ? | -- | ? | ? | -- | ? |
| #112-13 | A | -- | -- | -- | ? | -- | ? | ? | ? | ? |
| #117-03 | A | -- | -- | + | ? | + | + | + | + | ? |
| #121-03 | A | -- | + | ? | + | + | ? | ? | + | ? |

**Table 6-7.** Continued (2)

| Individual | Status** | Markers[a] (Centromere---Telomere) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | D16S423* AFM249 | D16S63* AC327 | D16S80* 24-1 | D16453 AC301G12 | D16S523 AC13H1 | D16S246* 218EP6 | D16S94 ACVK5 | D16S283* SM7 | D16S291 AC2.5 |
| #121-06 | A | -- | -- | ? | -- | -- | ? | ? | + | ? |
| #145-11 | A | -- | -- | -- | -- | -- | ? | ? | ? | ? |
| #F2-9 | N | ? | + | -- | -- | -- | ? | ? | + | + |
| #F3-13 | N | ? | + | ? | -- | -- | -- | ? | -- | -- |
| #F3-16 | N | ? | + | ? | + | + | -- | ? | -- | -- |
| #F11-69 | N | ? | -- | ? | -- | -- | ? | ? | + | + |
| #F14-59 | A | ? | + | ? | ? | ? | -- | ? | -- | ? |
| #F14-60 (maternal) | A | ? | -- | ? | ? | ? | + | ? | + | ? |
| #F14-60 (paternal) | A | ? | -- | -- | -- | + | ? | ? | ? | + |
| #F14-61 | N | ? | -- | ? | ? | ? | + | ? | + | ? |

not informative at both sides of crossovers in these recombinants, but these recombinants defined the most probable localization of the *MEF* gene.

### 6.3.4. Homozygosity Detection

The figure 6-1 is a summary of genotype status in the affected individuals of 18 consanguineous families. This detection was based on the strategy of homozygosity mapping (Lander and Botstein 1987) which relies on the fact that, in consanguineous families, children with rare recessive diseases usually inherit both copies of the disease gene, along with markers on adjacent chromosomal intervals, from a common ancestor. The homozygosity at *D16S246* (218EP6) locus in all the affected individuals except 72-05 was observed. This favours the likelihood that the location of the *MEF* gene is between 218EP6 and AC13H1. The heterozygosity at AC13H1 locus found in 10 affected individuals (except 72-05) suggests that the *MEF* gene might be located closer to 218EP6 locus than to AC13H1 locus. The individual 72-05 was heterozygous at AC13H1, 218EP6, ACVK5 and AC2.5. The heterozygosity at these loci is due to the fact that this individual's mother was affected, therefore, in this family there are three affected chromosomes (two maternal and one paternal) in the parents which transmitted this autosomal recessive disorder to their offspring.

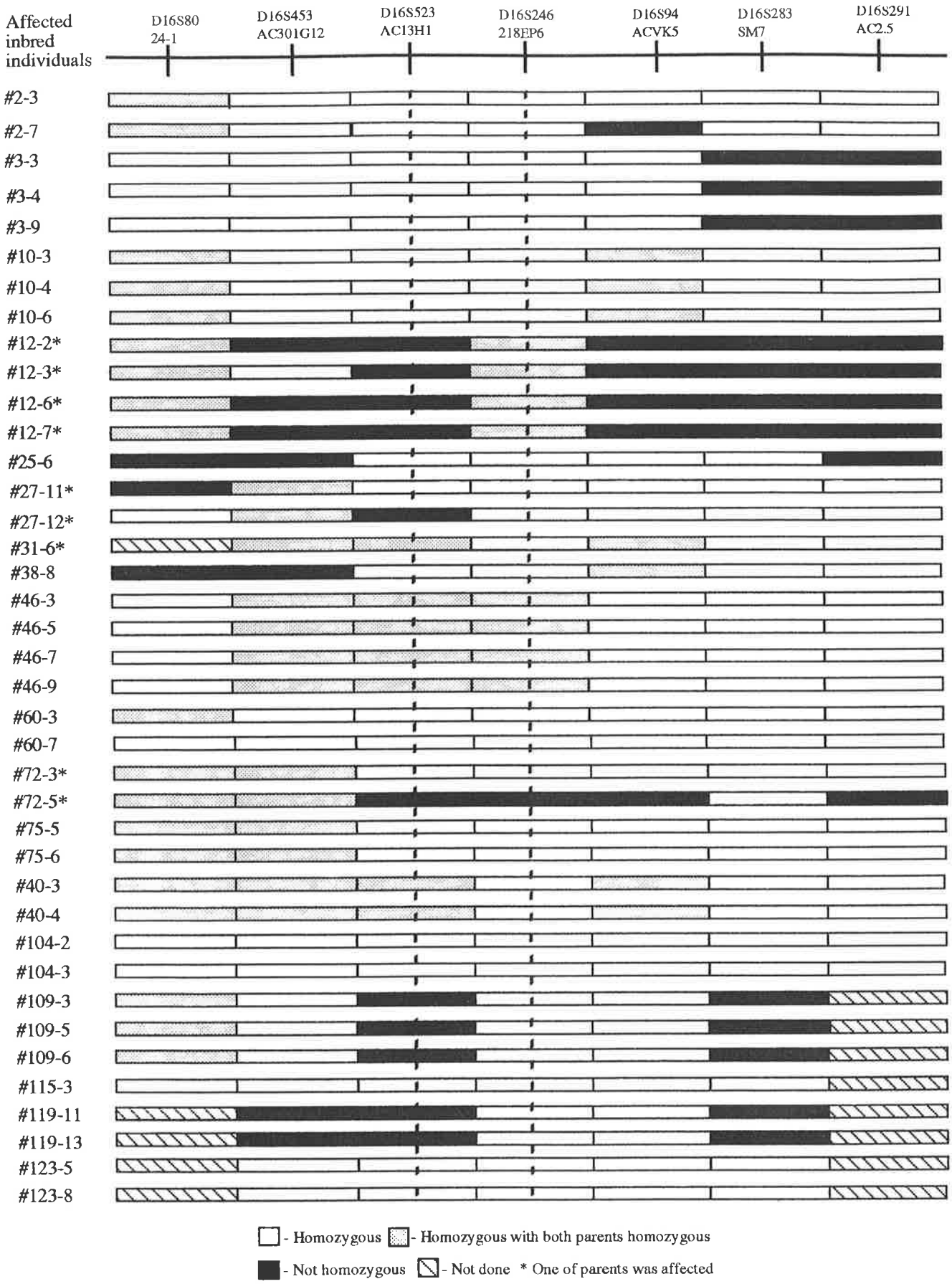



**Figure 6-1.** Genotype Status of Affected Individuals for 7 Loci in Consanguineous Families.

### 6.3.5. Linkage Analysis

#### 6.3.5.1. Two-Point Linkage Analysis

Two-point linkage analyses of three (AC)n marker loci (*D16S94, D16S523* and *D16S453*) and the *MEF* locus are summarized in Table 6-8. The lod scores showed these markers were tightly linked to *MEF* locus. The maximal lod scores for *D16S94, D16S523* and *D16S453* were 11.56 at $\theta = 0.04$, 24.29 at $\theta = 0.04$ and 18.96 at $\theta = 0.02$ (sex-averaged), respectively. The lod scores for *D16S291* (AC2.5) listed in Table 6-9 was from the previous publication (Aksentijevich et al., 1993b).

#### 6.3.5.2. Multipoint Linkage Analysis

The results of multipoint linkage analysis were presented in Table 6-10. The maximal lod score was 39.05 at $\theta = 0.008$, which placed the *MEF* gene locus between *D16S246* (218EP6) and *D16S523* (13H1). This confirmed the results from recombinants and homozygosity mapping.

## 6.4. DISCUSSION

### 6.4.1. Allele Association

Significant allelic association was only found between *D16S523* (AC13H1) and *MEF* in 14 Moroccan Jewish families. Although recombination events have

**Table 6-8.** Pairwise Lod Scores between MEF and 16p13.3 Markers (*D16S94, D16S523* and *D16S453* )

| Locus | θ (sex-averaged) | | | | | | | Maximal Lod Scores | Recombination Fraction (θ) Sex-Averaged |
|---|---|---|---|---|---|---|---|---|---|
| | 0.0 | 0.01 | 0.05 | 0.10 | 0.20 | 0.30 | 0.40 | | |
| D16S94 (ACVK5) | 8.07 | 10.63 | 11.52 | 10.61 | 7.55 | 4.15 | 1.29 | 11.56 | 0.04 |
| D16S523 (AC13H1) | 17.72 | 22.72 | 24.11 | 21.86 | 15.05 | 7.93 | 2.28 | 24.29 | 0.04 |
| D16S453 (AC301G12) | 17.11 | 18.66 | 18.48 | 16.38 | 11.04 | 5.78 | 1.69 | 18.96 | 0.02 |

**Table 6-9.** Pairwise Lod Scores between MEF and 16p13.3 Marker*D16S291* (Aksentijevich et al. 1993b)

| Locus | θ (sex-averaged) | | | | | | Maximal Lod Scores | θf | θm |
|---|---|---|---|---|---|---|---|---|---|
| | 0.0 | 0.01 | 0.02 | 0.05 | 0.10 | 0.15 | | | |
| D16S291 (AC2.5) | 11.52 | 15.83 | 16.71 | 17.10 | 16.10 | 13.80 | 17.16 | 0.03 | 0.04 |

**Table 6-10.** Results of Multipoint Linkage Analysis Using LINKMAP

| Recombination Fraction | 0.5 | 0.4 | 0.3 | 0.2 | 0.1 | 0 | 0.004 | **0.008** | 0.012 | 0.016 | 0.02 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | D16S246 | | | | | D16S523 | | | | | |
| Multipoint Lod Scores | 0 | 3.54 | 11.89 | 22.28 | 32.57 | 37.28 | 38.74 | **39.05** | 38.97 | 38.55 | 37.08 | 33.31 | 22.72 | 12.13 | 3.6 | 0 |
| Kosambi Distance (cM) | - | -55 | -35 | -21 | -10 | 0 | 0.4 | **0.8** | 1.2 | 1.6 | 2.0 | 12 | 23 | 37 | 57 | |

identified *D16S523* as the closest centromeric marker to *MEF*, there was no significant allelic association with this locus among 16 non-Moroccan Jewish families and 13 Armenian families. This is most likely because the 83bp *D16S523* allele frequency on FMF carrier chromosomes is relatively higher than on non-carrier chromosomes in Moroccan Jews. In contrast, the 83bp *D16S523* allele frequency on FMF carrier chromosomes is lower than on non-carrier chromosomes in non-Moroccans and Armenians. It is also seen that the 83bp *D16S523* allele is the most common allele in the normal population (CEPH data, Table 6-3). The further centromeric marker *D16S453* (AC301G12) was also analysed. Similarly, although the frequency of the 125bp *D16S453* allele in FMF carrier chromosomes is the highest, this allele also has the highest frequency in non-carrier chromosomes (Table 6-4 to 6-6) and in the normal population (Table 6-3). This marker is also relatively uninformative. Therefore, testing linkage disequilibrium in this study, which showed the *MEF* gene is associated with a common allele of the flanking marker, is not very powerful for supporting the localization of the *MEF* gene. However, the significant allelic association in the Moroccan Jewish subpopulation may provide evidence for a founder effect in this ethnic group.

The allelic association between two distal markers [*D16S291* (AC2.5) and *D16S94* (ACVK5)] and *MEF* was observed to show the same results (Aksentijevich et al., 1993c). The 86bp *D16S94* allele was not significantly associated with *MEF* in Moroccan Jewish families, although *D16S94* was the

closest of the three markers (*D16S94, D16S283, D16S291*) to *MEF*. In contrast, the 93bp *D16S283* (SM7) allele and the 162bp *D16S291* (AC2.5) allele were not the most common alleles in non-carrier chromosomes (Aksentijevich et al., 1993c) or in the normal population (CEPH data, Thompson et al., 1992; Harris et al., 1991), and these two markers showed significant allelic association with *MEF* (Aksentijevich et al., 1993c).

### 6.4.2. Localization of the *MEF* Gene

The genetic background map of 16p13.3 region based on STR markers was constructed by linkage analysis of genotype data from the CEPH reference families (Table 4-6 and Figure 4-6). The order of loci and the distances between loci are as follows:16pter-*D16S291*(AC2.5)$\underline{^{1.1cM}}$*D16S283*(SM7)$\underline{^{0.4cM}}$*D16S94*(ACVK5) $\underline{^{1.6cM}}$*D16S523*(AC13H1)$\underline{^{1.8cM}}$*D16S453*(AC301G12)/*D16S423*(AFM249)--cen. The RFLP marker *D16S246* (218EP6) was physically mapped in the same breakpoint interval as *D16S291, D16S283* and *D16S94* (data not show) and the genetic location on the CEPH consortium map is between *D16S94* and *D16S523* (personal communication, H. Kozman, 1994). The odds to support the order at locus intervals *D16S283* and *D16S94*, *D16S523* and *D16S453/D16S423* were not greater than 1,000 : 1 (Table 4-6). This is possibly due to insufficient information forthcoming from genotyping of only the 8 largest CEPH families for some of these markers (Weissenbach et al., 1992). The obvious example is the ordering of

*D16S423*. There was no recombination observed between *D16S453* and *D16S423* on the CEPH background map, but recombination events were observed between *D16S453* and *D16S423, D16S80* or *D16S63* in FMF families (Table 6-7) indicating that *D16S423* is centromeric to *D16S453*. The final locus order was determined from the CEPH background map (Table 4-6), previous linkage analysis (Aksentijevich et al., 1993b) and observed recombination events in FMF families (Table 6-7):

16pter--*D16S291* (AC2.5) -- *D16S283* (SM7) -- *D16S94* (ACVK5) -- *D16S246* (218EP6) --*D16S523* (AC13H1) -- *D16S453* (AC301G12) -- *D16S80* (24.1) -- *D16S63* (AC327) -- *D16S423* (AFM249) -- centromere

In this study, four (AC)n repeat markers (*D16S291, D16S94, D16S523* and *D16S453*) were demonstrated to be tightly linked to *MEF* by two-point linkage analysis (Table 6-8 and 6-9).

From 35 observed recombination events in 62 FMF families, only one recombination event in the normal individual #F3-16 significantly reduced the regional localization and redefined the centromeric boundary at *D16S523*(AC13H1) and telomeric boundary at *D16S246* (218EP6) because both markers were informative (Table 6-7). *D16S523* and *D16S246* were not fully informative to confirm the exact crossovers between these two markers in six recombinants (13-

10, 72-05, 112-13, 121-06, 145-11 and F14-59, Table 6-7), these recombinants, however, favour the localization of the *MEF* gene between *D16S246* and *D16S523*.

Recessive diseases in the offspring of consanguineous parents usually arise because the affected individuals have two defective gene copies from the common ancestor and therefore closely linked flanking genetic markers should also be homozygous. This strategy for chromosomal localization of recessive disease genes was termed "homozygosity mapping" (Lander and Botstein, 1987). The gene responsible for alkaptonuria has been localized to 3q2 by homozygosity mapping (Pollak et al., 1993). For localization of *MEF*, a homozygosity mapping approach has been used in a previous study (Aksentijevich et al., 1993b). In this study, homozygosity in affected offspring of consanguineous families supported the localization of the *MEF* gene. All of the affected individuals were homozygous at the *D16S246* (218EP6) locus (Figure 6-1). The other loci were shown to be heterozygous (Figure 6-1). It is known that the gene frequency for *MEF* is extraordinarily high (0.045) in this study population. Due to this high frequency, it is possible that affected individuals were not necessarily homozygous at the closest centromeric locus AC13H1 by descent of a different allele inherited with MEF from a married-in parent. The higher heterozygosity (0.62) of AC13H1 than 218EP6 (0.46) in FMF families also suggests this possibility. The homozygosity at 218EP6 locus, nevertheless, significantly supports the possibility that *MEF* is located between AC13H1 and 218EP6, and possibly closer to 218EP6 than AC13H1.

The multipoint linkage analysis (Table 6-10) placed *MEF* centromeric to *D16S246* (218EP6) and telomeric to *D16S523* (AC13H1). This result was consistent with the localization of *MEF* between these two markers by recombinant analysis and homozygosity mapping. Although the linkage analysis showed that the maximal multipoint lod score 39.05 was 0.8 cM centromeric to *D16S246*, whether the *MEF* gene locus is closer to *D16S246* than *D16S523* can not be determined because the multipoint lod scores between these two markers, with recombination fractions from 0.0 to 0.02, are very similar. The distances between *MEF* and these two markers, therefore, will be determined using physical mapping from the next step of positional cloning of *MEF* by developing a YAC contig across this region.

In previous studies, 1) the *MEF* gene was localized between *D16S94* (ACVK5) and *D16S80* (24.1) in a genetic interval of about 9 cM (Aksentijevich et al., 1993b), 2) the MEF gene was located between *D16S283* (SM7) and *D16S80* (24.1) at a distance of approximately 3 cM centromeric from SM7 (Fischel-Ghodsian et al., 1993). At present, there is no genetic linkage map available which includes both STRs and RFLPs in 16p13.3 region. The distance flanking *MEF* can only be estimated at this time by the CEPH background map based on STRs (Table 4-6). The distance between *D16S94* (ACVK5) and *D16S523* (AC13H1) is 1.6 cM on the CEPH map. The RFLP marker *D16S246* (218EP6) is centromeric to *D16S94*. The *MEF* gene, thus, is likely to be located within a genetic interval of less than 1.6 cM.

The results from recombination study, homozygosity mapping and two-point and multipoint linkage analysis positively further refined the localization of the *MEF* gene. The narrowed interval and closer flanking markers allow positional cloning to be more feasible and efficient, and in the meantime provide the possibility of more accurate preclinical diagnosis.

## 6.5. CONCLUSION

To refine localization of the gene responsible for FMF, four (AC)n repeat markers (*D16S291, D16S94, D16S523* and *D16S453*) were selected from the CEPH background map of chromosome 16 based on STR markers. Sixty-two FMF families were genotyped using these 4 markers. Allelic association, two-point and multipoint linkage analysis, recombination analysis and homozygosity mapping were used in this study. The four markers were tightly linked to FMF by two-point linkage analysis. The observed recombination events, homozygosity mapping and multipoint linkage analysis defined the centromeric boundary at *D16S523* (13H1) and telomeric boundary at *D16S246* (218EP6). The genetic distance between these two markers was estimated to be less than 1.6 cM based on CEPH background map. This refined localization of the *MEF* gene will facilitate the positional cloning of this gene and aid in the preclinical diagnosis of patients prior to isolation of the gene.

# CHAPTER 6

## CONCLUSION

The aims of this project were achieved. A high-resolution cytogenetic-based physical map and genetic linkage map of human chromosome 16 was developed based on 79 STR markers. Two approaches were used to isolate STR markers (Chapter 3). Initially the random isolation approach was exploited to generate STR markers from the chromosome 16 cosmid library (Stallings et al., 1990). After (AC)n repeats isolated from the random approach were mapped to 9 intervals of chromosome 16 defined by breakpoints in somatic cell hybrids (chapter 4), a second targetted approach was used to generate additional repeat markers with which to fill in the deficient intervals or to isolate more repeat markers in the regions of particular interest [fragile sites *FRA16A* and *FRA16B*, disease gene regions of adult autosomal polycystic kidney disease (PKD1), familial mediterranean fever (FMF) and Batten disease (CLN3)]. This second approach involved the isolation of STR markers from cosmid clones which had been previously mapped to specific regions of chromosome 16 (Callen et al., 1992; Stallings et al., 1992).

A total of **32** (AC)n repeat markers were isolated from these two approaches. Twenty-two (AC)n repeat markers had heterozygosity greater than 0.68 (chapter 3). The **32** (AC)n repeat markers and **20** (AC)n repeat markers isolated by other members of the candidate's laboratory, together with **33** (AC)n repeat markers from other laboratories were physically mapped on the high-resolution cytogenetic map of chromosome 16 using the hybrid panel (chapter 4). Seventy-nine STR markers

were used for construction of a genetic linkage map of human chromosome 16 (chapter 4, Shen et al., submitted). Of these 27 markers were generated by the candidate.

This STR-based genetic linkage map of human chromosome 16 covers the entire length of chromosome 16 (chapter 4). The length of the map is in remarkable agreement with the previously published genetic and chiasma maps of chromosome 16 (Shen et al, submitted). The median distance and the average distance between markers on the framework map is 2.6 and 3.2 cM, respectively. This resolution has reached one goal of the Human Genome Project which aimed to construct genetic linkage maps of all human chromosomes with highly polymorphic markers spaced an average of 2-5 centimorgans apart (Jordan, 1992).

The framework and comprehensive maps were anchored to the high-resolution cytogenetic map (chapter 4), which was divided into 66 breakpoint intervals (on average 1.5 Mb per interval) by a panel of 67 hybrids. The cytogenetic map is one of the most detailed maps available for any of the autosomes (Callen et al., 1992). It is apparent that the combination of genetic linkage analysis and physical mapping can be extremely helpful in resolving locus order at the resolution of the comprehensive map.

These integrated genetic and physical maps of human chromosome 16 (Shen et al. submitted) provide an efficient means for regional localization of genetic disorders on chromosome 16, for detection of loss of heterozygosity in cancers and genomic

imprinting of chromosomes in inherited disorders, for evaluation of linkage disequilibrium and disease causing mutations, and for analysis of multifactorial diseases.

For positional cloning of disease genes, one strategy is the identification of their location in the genome by genetic linkage mapping as the first step, followed by the refined localization of the disease gene between flanking genetic markers and then isolation of a contig of large genomic fragments across the region (between the flanking genetic markers) can be developed. In this study, the genetic and physical maps of chromosome 16 based on STR markers provide the basis for refining the localization of the *MEF* gene. Four STR markers (*D16S291, D16S94, D16S523* and *D16S453*), which were genetically and physically mapped in chromosome 16p13.3 region, were selected to genotype 62 FMF families. The linkage analysis was carried out using the computer program LINKAGE (version 5.1). The lod scores from two-point linkage analysis showed these 4 markers were significantly linked to *MEF*. The observed recombinants, homozygosity mapping and multipoint linkage analysis refined the localization of *MEF* between *D16S523* and *D16S246*. Based on the constructed CEPH background map, the *MEF* gene is located in a genetic interval of less than 1.6 cM. This refined localization provides the basis for the next step of positional cloning of *MEF* by developing a YAC contig across this region, and also provides closer flanking markers for preclinical diagnosis.

# References

Aaltonen LA, Peltomaki P, Leach FS, Sistonen P, Pylkkanen L, Mecklin J-P, Jarvinen H, Powell SM, Jen J, Hamilton SR, Petersen GM, Kinzler KW, Vogelstein B and de la Chapelle A (1993) Clues to the pathogenesis of familial colorectal cancer. *Science* 260: 812-816

Aksentijevich I, Gruberg L, Balow Jr. JE, Dean M, Pras M and Kastner DL (1991) Linkage analysis in familial Mediterranean fever (FMF). *Am J Hum Genet Suppl* 49: 335

Aksentijevich I, Gruberg L, Balow Jr. JE, Kovo M, Gazit E, Dean M, Pras M and Kastner DL (1993a) Evidence for linkage of the gene causing familial Mediterranean fever to chromosome 17q in non-Ashkenazi Jewish families: second locus or type I error. *Hum Genet* 91: 527-534

Aksentijevich I, Pras E, Gruberg L, Shen Y, Holman K, Helling S, Prosin L, Sutherland GR, Richards RI, Ramsurg M, Dean M, Pras M, Amos CI and Kastner DL (1993b) Refined mapping of the gene causing familial Mediterranean fever by linkage and homozygosity studies. *Am J Hum Genet* 53: 451-461

Aksentijevich I, Pras E, Gruberg L, Shen Y, Holman K, Helling S, Prosin L, Sutherland GR, Richards RI, Dean M, Pras M and Kastner DL (1993c) Familial Mediterranean fever in Moroccan Jews: demonstration of a founder effect by extended haplotype analysis. *Am J Hum Genet* 53: 644-651

Baker E, Guo X-H, Orsborn AM, Sutherland GR, Callen DF, Hopwood JL and Morris PC (1993) The Morquio A syndrome (Mucopolysaccharidosis IVA) gene maps to 16q24.3. *Am J Hum Genet* 52: 96-98

Beckmann JS and Weber JL (1992) Survey of human and rat microsatellites. *Genomics* 12:627-631

Beckmann JS, Tomfohrde J, Barnes RI, Williams M, Broux O, Richard I, Weissenbach J and Bowcock AM (1993) A linkage map of human chromosome 15 with an average resolution of 2 cM and containing 55 polymorphic microsatellites. *Hum Mol Genet* 2: 2019-2030

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

Bowcock A, Osborne-Lawrence S, Barnes R, Chakravarti A, Washington S and Dunn C (1993) Microsatellite polymorphism linkage map of human chromosome 13q. *Genomics* 15: 376-386

Braaten DC, Thomas JR, Little RD, Dickson KR, Goldberg I, Schlessinger D, Ciccodicola A and D'Urso M (1988) Locations and contexts of sequences that hybridize to poly(dG-dT)·(dC-dA) in mammalian ribosomal DNAs and two X-linked genes. *Nucl Acid Res* 16: 865-888

Breuning MH, Dauwerse HG, Fugazza G, Saris JJ, Spruit L, Wijnen H, Tommerup N, van der Hagen CB, Imazumi K, Kuroki Y, van der Boogaard M-J, de Pater JM, Mariman ECM, Hamel BCJ, Himmelbauer H, Frischauf A-M, Stallings RL, Beverstok GC, van Ommen G-JB and Hennekam RCM (1993) Rubinstein-Taybi syndrome caused by submicroscopic deletions within 16p13.3. *Am J Hum Genet* 52: 249-254

Brook JD, McCurrach ME, Harley HG, Bruckler AJ, Church D, Aburatani H, Hunter K, Stanton VP, Thirion JP, Hudson T, Sohn R, Lemelman B, Snell RG, Rundle SA, Crow S, Davies J, Shelbourne P, Buxton J, Jones C, Juvonen V, Johnson K, Harper PS, Shaw DJ and Housman DE (1992) Molecular basis of myotonic dystrophy: Expansion of a trinucleotide (CTG) repeat at the 3' end of a transcript encoding a protein kinase family member. *Cell* 68: 799-808

Brooks-Wilson AR, Smailus DE, Weier H-UG and Goodfellow PJ (1992) Human repeat element-mediated PCR: cloning and mapping of chromosome 10 DNA markers. *Genomics* 13:409-414

Buetow KH (1991) Influence of aberrant observations on high-resolution linkage analysis outcomes. *Am J Hum Genet* 49: 985-994

Buetow KH, Duggan D, Yang B, Ludwigsen S, Puck J, Porter J, Budarf M, Spielman R and Emanuel BS (1993) A microsatellite-based multipoint index map of human chromosome 22. *Genomics* 18: 329-339

Bullock WO, Fernandez JM and Short JM (1987) XL-1 Blue: A high effiency plasmid transforming recA *Escherchia coli* strain with β-galactosidase selection. *Biotechniques* 5: 376

Callen DF (1986) A mouse-human hybrid cell panel for mapping human chromosome 16. *Ann Genet* 29: 235-239

Callen DF, Baker E, Eyre HJ and Lane SA (1990) An expanded mouse-human hybrid cell panel for mapping human chromosome 16. *Ann Genet* 4: 190-195

Callen DF, Doggett NA, Stallings RL, Chen LZ, Whitmore SA, Lane SA, Nancarrow JK, Apostolou S, Thompson AD, Lapsys NM, Eyre HJ, Baker EG, Shen Y, Holman K, Phillips H, Richards RI and Sutherland GR (1992) High-resolution cytogenetic-based physical map of human chromosome 16. *Genomics* 13:1178-1185

Callen DF, Thompson AD, Shen Y, Phillips HA, Richards RI, Mulley JC and Sutherland GR (1992) Incidence and origin of "Null" alleles in the (AC)n microsatellite markers from chromosome 16. *Am J Hum Genet* 52: 922-927

Clemens PR, Frenwick RG, Chamberlain JS, Gibbs RA, de Andrade M, Chakraborty R and Caskey CT (1991) Carrier detection and prenatal diagnosis in Duchenne and Becker muscular dystrophy families, using dinucleotide repeat polymorphisms. *Am J Hum Genet* 49: 951-960

Dausset J, Cann H, Cohen D, Lathrop M, Lalouel J-M and White R (1990) Centre d'Etude du Polymorphisme Humain (CEPH): collaborative genetic mapping of the human genome. *Genomics* 6: 575-577

Davies KE, Pearson PL, Harper PS, Murray JM, O'Brien T, Sarfarazi M and Williamson R (1983) Linkage analysis of two cloned sequences flanking the Duchenne muscular dystrophy locus on the short arm of the human X chromosome. *Nucl Acids Res* 11: 2303-2312

Dawson ES, Shaikh S, Weber JL, Wang Z, Weissenbach J, Powell JF and Gill M (1993) A continuous linkage map of 22 short tandem repeat polymorphisms on human chromosome 12. *Genomics* 17: 245-248

Decker RA, Moore J, Ponder B and Weber JL (1992) Linkage mapping of human chromosome 10 microsatellite polymorphisms. *Genomics* 12: 604-606

Devlin B, Risch N and Roeder K (1991) Estimation of allele frequencies for VNTR loci. *Am J Hum Genet* 48: 662-676

Dietrich W, Katz H, Lincoln SE, Shin H-S, Friedman J, Dracopoli NC and Lander ES (1992) A genetic map of the mouse suitable for typing intraspecific crosses. *Genetics* 131: 423-447

Divis MM, Calame K, Early PW, Livant DL, Joho R, Weissman IL and Hood L (1980) An immunoglobulin heavy-chain gene is formed by at least two recombinational events. *Nature* 283: 733-739

Doggett NA (1992) The mapping of chromosome 16. *Los Alamos Science* (20): 182-217

Donis-Keller H, Green P, Helms C, Cartinhour S, Weiffenbach B, Stephens K, Keith TP, Bowden DW, Smith DR, Lander ES, Botstein D, Akots G, Rediker KS, Gravius T, Brown VA, Rising MB, Parker C, Powers JA, Watt DE, Kauffman ER, Bricker A, Phipps P, Muller-Kahle H, Fulton TR, Ng S, Schumm JW, Braman JC,

Knowlton RG, Barker DF, Crooks SM, Lincoln SE, Daly MJ and Abrahamson J (1987) A genetic linkage map of the human genome. *Cell* 51: 319-337

Drwinga HL, Toji LH, Kim CH, Greene AE and Mulivor RA (1993) NIGMS human-rodent somatic cell hybrid mapping panel 1 and 2. *Genomics* 16: 311-314

Economou EP, Bergen AW, Warren AC and Antonarakis SE (1990) The polydeoxyadenylate tract of Alu repetitive elements is polymorphic in the human genome. *Proc Natl Acad Sci USA* 87: 2951-2954

Edwards A, Civitello A, Hammond HA and Caskey CT (1991) DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *Am J Hum Genet* 49: 746-756

Edwards A, Hammond HA, Jin L, Caskey CT and Chakraborty R (1992) Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. *Genomics* 12: 241-253

Emi M, Fujiwara Y and Nakamura Y (1993) A primary genetic linkage map of 14 polymorphic loci for the short arm of human chromosome 8. *Genomics* 15: 530-534

Engelstein M, Hudson TL, Lane JM, Lee MK, Leberone B, Landes GM, Peltonen L, Weber JL and Dracopoli NC (1993) A PCR-based linkage map of human chromosome 1. *Genomics* 15: 251-258

Fischel-Ghodsian N, Bu X-D, Prezant TR, Oeztas S, Huang Z-S, Bohlman MC, Rotter JI and Shohat M (1993) Regional mapping of the gene for familial Mediterranean fever on human chromosome 16p13. *Am J Med Genet* 46: 689-693

Froguel P, Vaxillaire M, Sun F, Velho G, Zouali H, Butel MO, Lesage S, Vionnet N, Clement K, Fougerousse F, Tanizawa Y, Weissenbach J, Beckmann JS, Lathrop GM, Passa P, Permutt MA and Cohen D (1992) Close linkage of glucokinase locus on chromosome 7p to early-onset non-insulin-dependent diabetes mellitus. *Nature* 356: 162-164

Goff DJ, Galvin K, Katz H, Westerfield M, Lander ES and Tabin CJ (1992) Identification of polymorphic simple sequence repeats in the genome of the Zebrafish. *Genomics* 14: 200-202

Goltsov AA, Eisensmith RC, Naughton ER, Jin L, Chakraborty R and Woo SLC (1993) A single polymorphic STR system in the human phenylalanine hydroxylase gene permits rapid prenatal diagnosis and carrier screening for phenylketonuria. *Hum Mol Genet* 2: 577-581

Green P, Fall K and Crooks S (1990) Documentation from CRI-MAP, version 2.4 (3/26/90).

Gross DS, Huang SY and Garrard WT (1985) Chromatin structure of the potential Z-forming sequence (dT-dG)n.(dC-dA)n evidence for an "alternating-B" conformation. *Med Biol* 183: 251-265

Gross DS and Garrard WT (1986) The ubiquitous potential Z-forming sequence of eukaryotes, (dT-dG)n.(dC-dA)n, is not detectable in the genomes of eubacteria, archaebacteria, or mitochondria. *Mol Cel Biol* 6: 3010- 3013

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

Hamada H and Kakunaga T (1982a) Potential Z-DNA forming sequences are highly dispersed in the human genome. *Nature* 298: 396-398

Hamada H, Petrino MG and Kakunaga T (1982b) A novel repeated element with Z-DNA-forming potential is widely found in evolutionarily diverse eukaryotic genomes. *Proc Natl Acad Sci USA* 79: 6465-6469

Hamada H, Seidman M, Howard BH and Gorman CM (1984) Enhanced gene expression by the poly(dT-dG).poly(dC-dA) sequence. *Mol Cell Biol* 4: 2622-2630

Harris P, Thomas S, Ratcliffe PJ, Breuning MH, Coto E and Lopez-Larrea C (1991) Rapid genetic analysis of families with polycystic kidney disease 1 by means of a microsatellite marker. *Lancet* 338: 1484-1487

Hazan J, Dubay C, Pankowiak M-P, Becuwe N and Weissenbach J (1992) A genetic linkage map of human chromosome 20 composed entirely of microsatellite markers. *Genomics* 12: 183-1

Hetzel J (1993) Livestock genome research on the march. *Nature Genetics* 4: 327-328

Hudson TJ, Engelstein M, Lee MK, Ho EC, Rubenfield MJ, Adams CP, Housman DE and Dracopoli NC (1992) Isolation and chromosomal assignment of 100 highly informative human simple sequence repeat polymorphisms. *Genomics* 13: 622-629

Hyland VJ, Suthers GK, Friend D, MacKinnon RN, Callen DF, Breuning MH, Keith MH, Brown T, Phipps VA and Sutherland GR (1990) Probe, VK5B, is located in the same interval as the autosomal dominant adult polycystic kidney disease locus, PKD 1. *Hum Genet* 84: 286-288

Jordan E (1992) The human genome project: where did it come from, where is it going? *Am J Hum Genet* 51: 1-6

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

Jodice C, Frontali M, Persichetti F, Novelletto A, Pandolfo M, Spadaro M, Giunti P, Schinaia G, Lulli P, Malaspina P, Plasmati R, Tola R, Antonelli A, Donato SD, Morocutti C, Weissenbach J, Cann HM and Terrenato L (1993) The gene for spinal cerebellar ataxia 1 (SCA1) is flanked by two closely linked highly polymorphic microsatellite loci. *Hum Mol Genet* 2: 1383-1387

Julier C, Nakamura Y, Lathrop M, O'Connell P, Leppert M, Mohandas T, Lalouel J-M and White R (1990) A primary map of 24 loci on human chromosome 16. *Genomics* 6: 419-427

Kaback DB, de Steensma HY and de Jonge P (1989) Enhanced meiotic recombination on the smallest chromosome of *Saccharomyces cerevisiae*. *Proc Natl Acad Sci USA* 86: 3694-3698

Kandt RS, Haines JL, Smith M, Northrup H, Gardner RJM, Short MP, Dumars K, Roach ES, Steingold S, Wall S, Blanton SH, Flodman P, Kwiatkowski DJ, Jewell A, Weber JL, Roses AD and Periccak-Vance MA (1992) Linkage of an important gene locus for tuberous sclerosis to a chromosome 16 marker for polycystic kidney disease. *Nature Genetics* 2: 37-41

Keats BJ, Sherman SL, Morton NE, Robson EB, Buetow KH, Cartwright PE, Chakravarti A, Francke U, Green PP and Ott J (1991) Guideline for human linkage maps: an international system for human linkage maps. *Genomics* 9: 557-560

Keith TP, Green P, Reeders ST, Brown VA, Phipps P, Bricker A, Falls K, Rediker KS, Powers JA, Hogan C, Nelson C, Knowlton R and Donis-Keller H (1990) Genetic linkage map of 46 DNA markers on human chromosome 16. *Proc Natl Acad Sci USA* 87: 5754-5758

Kogan SC, Doherty BS, Gitschier J (1987) An improved method for prenatal diagnosis of genetic diseases by analysis of amplified DNA sequences. *N Eng J Med* 317: 985-990

Konradi C, Ozelius L, Yan W, Gusella JF and Breakefield XO (1991) Dinucleotide repeat polymorphism (D16S285) on human chromosome 16. *Nucl Acid Res* 19: 5449

Kozman HM, Phillips HA, Callen DF, Sutherland GR and Mulley JC (1993) Integration of the cytogenetic and genetic linkage maps of human chromosome 16 using 50 physical intervals and 50 polymorphic loci. *Cytogenet and Cell Genet* 62: 194-198

Kremer EJ, Pritchard M, Lynch M, Yu S, Holman K, Baker E, Warren ST, Schlessinger D, Sutherland GR and Richards RI (1991) Mapping of DNA instability at the fragile X to a trinucleotide repeat p(CCG)n. *Science* 252: 1711-1714

Kwiatkowski DJ, Hinske EP, Weimer K, Ozelius L, Gusella JF and Haines J (1992) Construction of a GT polymorphism map of human 9q. *Genomics* 12: 229-240

Lander ES and Botstein D (1987) Homozygosity mapping : a way to map human recessive traits with the DNA of inbred children. *Science* 236: 1567-1570

Lander ES and Green P (1987) Construction of multilocus genetic linkage maps in humans. *Proc Natl Acad Sci, USA* 84: 2363-2367

Lander ES, Green P, Abrahamaon J, Barlow A, Daly MJ, Lincoln SE and Newburg L (1987) Mapmaker: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* 1: 174-181

La Spada AR, Wilson EM, Lubahn DB, Harding AE and Fischbeck H (1991) Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. *Nature* 352:77-79

Lathorp G and Lalouel JM (1984) Easy calculation of lod scores and genetic risks on small computers. *Am J Hum Genet* 36: 460-465

Lathorp G, Lalouel JM, Julier C and Ott J (1984) Strategies for multilocus linkage analysis in human. *Proc Natl Acad Aci USA* 81: 3443-3446

Lindblom A, Tannergard P, Werelius B and Nordenskjold M (1993) Genetic mapping of a second locus predisposing to hereditary non-polyposis colon cancer. *Nature Genetics* 5: 279-282

Litt M and Luty JA (1989) A hypervariable microsatellite revealed by *in vitro* amplification of a dinucleotide repeat within the cardiac muscle actin gene. *Am J Hum Genet* 44: 397-401

Litt M, Kramer P, Hauge XY, Weber JL, Wang Z, Wilkie PJ, Holt MS, Mishra S, Donis-Keller H, Warnich L, Retief AE, Jones C and Weissenbach J (1993) A microsatellite-based index map of human chromosome 11. *Hum Mol Genet* 2: 909-913

MacDonald ME, The Huntington's Disease Collaborative Research Group (1993) A novel gene containing a trinucleotide repeat that is expanded and unstable in Huntington's disease chromosomes. *Cell* 72: 971-983

Maniatis T, Fritsh EF and Sanbrook J (1982) Molecular Cloning: A Laboratory Manual. First Edition, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York

Maniatis T, Fritsh EF and Sanbrook J (1989) Molecular Cloning: A Laboratory Manual. Second Edition, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York

McInnis M, Chakravarti A, Blaschak J, Petersen MB, Sharma V, Avramopoulos D, Blouin J-L, Konig U, Brahe C, Matise TC, Warren AC, Talbot CC, Broeckhoven CV, Litt M and Antonarakis SE (1993) A linkage map of human chromosome 21: 43 PCR markers at average intervals of 2.5 cM. *Genomics* 16: 562-571

McKusick VA (1992) Autosomal recessive phenotypes. *In* Mendelian Inheritance In Man. Tenth Edition, pp.1518-1520, The John Hopskins University Press, Baltimore and London.

Meisfeld R, Krystal M and Arnhein N (1981) A member of a new repeated sequence family which is conserved throughout eucaryotic evolution is found between the human δ and β globin genes. *Nucl Acid Res* 9: 5931-5947

Melis R, Bradley P, Elsner T, Robertson M, Lawrence E, Gerken S, Albertsen H and White R (1993) Polymorphic SSR (Simple-Sequence-Repeat) markers for chromosome 20. *Genomics* 16: 56-62

Mills KA, Buetow KH, Xu Y, Weber JL, Altherr MR, Wasmuth JJ and Murray JC (1992) Genetic and physical maps of human chromosome 4 based on dinucleotide repeats. *Genomics* 14: 209-219

Mitchison HM, Thompson AD, Mulley JC, Kozman HM, Richards RI, Callen DF, Stallings RI, Doggett NA, Attwood J, McKay TR, Sutherland GR and Gardiner MR

(1993) Fine genetic mapping of the Batten disease locus (CLN3) by haplotype analysis and demonstration of allele association with chromosome 16p microsatellite loci. *Genomics* 16: 455-460

Morton NE, Lalouel J-M, Jackson JF, Currier RD and Yee S (1980) Linkage studies in spinocerebellar ataxia (SCA). *Am J Med Genet* 6: 251-257

Morton NE and Collins A (1990) Standard maps of chromosome 10. *Ann Hum Genet* 54: 235-251

Morton NE (1991) Parameters of the human genome. *Proc Natl Acad Sci, USA* 88: 7474-7476

Mulley JC and Sutherland GR (1993) Integrating maps of chromosome 16. *Current Opinion in Genetics and Development* 3: 425-431

Nakamura Y, Leppert M, O'Connell P, Wolff TH, Culver M, Martin C, Fujimoto E, Hoff M, Kumlin E and White R (1987) Variable number of tandem repaet (VNTR) markers for human mapping. *Science* 235: 1616-1622

NIH/CEPH Collaborative Mapping Group (1992) A comprehensive genetic linkage map of the human genome. *Science* 258: 67-162

Nordheim A and Rich A (1983) The sequence (dC-dA)n.(dG-dT) forms left-handed Z-DNA in negatively supercoiled plasmids. *Proc Natl Acad Sci USA* 80: 1821-1825

Orr HT, Chung M-Y, Banfi S, Kwiatkowski Jr. TJ, Servadio A, Beaudet AL, McCall AE, Duvick LA, Ranum LPW and Zoghbi HY (1993) Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1. *Nature Genetics* 4: 221-226

Ostrander EA, Sprague GF and Rine J (1993) Identification and characterization of dinucleotide repeat (CA)n markers for genetic mapping in dog. *Genomics* 16 : 207-213

Ott J (1992) Strategies for characterizing highly polymorphic markers in human gene mapping. *Am J Hum Genet* 51:283-290

Ozelius LJ, Kwiatkowski DJ, Schuback DE, Breakefield XO, Wexler NS, Gusella JF and Haines JL (1992) A genetic linkage map of human chromosome 9q. *Genomics* 14: 715-720

Pardue ML, Lowenhaupt K, Rich A and Nordheim A (1987) (dC-A)n.(dG-dT)n sequences have evolutionarily conserved chromosomal locations in *Drosophila* with implications for roles in chromosome structure and function. *EMBO J* 6: 1781-1789

Peltomaki P, Aaltonen LA, Sistonen P, Pylkkanen L, Mecklin J-P, Jarvinen H, Green JS, Jass JR, Weber JL, Leach FS, Petersen GM, Hamilton SR, de la Chapelle A and Vogelstein B (1993) Genetic mapping of a locus predisposing to human colorectal cancer. *Science* 260: 810-812

Petersen MB, Schinzel AA, Binkert F, Tranebjaerg L, Mikkelsen M, Collins FA and Economou EP (1991) Use of short sequence repeat DNA polymorphisms after PCR amplification to detect the parental origin of the additional chromosome 21 in Down syndrome. *Am J Hum Genet* 48: 65-71

Petersen MB, Antonarakis SE, Hassold TJ, Freeman SB, Sherman SL, Avramopoulos D and Mikkelsen M (1993) Paternal nondisjunction in trisomy 21: excess of male patients. *Hum Mol Genet* 2: 1691-1695

Petrukhin KE, Speer MC, Cayanis E, Bonaldo MF, Tantravahi U, Soares MB, Fischer SG, Warburton D, Gelliam TC and Ott J (1993) A microsatellite genetic linkage map of human chromosome 13. *Genomics* 15: 76-85

Phillips HA, Harris P, Richards RI, Sutherland GR and Mulley JC (1991) Dinucleotide repeat polymorphisms at the *D16S164, D16S168* and *D16S186* loci at 16q21-q22.1. *Nucl Acid Res* 19: 6964

Phillips HA, Hyland VJ, Holman K, Callen DF, Richards RI And Mulley JC (1991) Dinucleotide repeat polymorphism at *D16S287. Nucl Acid Res* 19: 6664

Phillips HA, Thompson AD, Kozman HM, Sutherland GR and Mulley JC (1993) A microsatellite marker within the duplicated D16S79A locus has a null allele : significance for linkage mapping. *Cytogenet Cell Genet* 64: 131-1342

Pollak MR, Chou Y-H W, Cerda JJ, Steinmann B, Du BJL, Seidman JG and Seidman CE (1993) Homozygosity mapping of the gene for alkaptonuria to chromosome 3q2. *Nature Genetics* 5: 201-204

Pras E, Aksentijevich I, Gruberg L, Balow Jr. JE, Prosen L, Dean M, Steinberg AD, Pras M and Kastner DL (1992) Mapping of a gene causing familial Mediterranean fever to the short arm of chromosome 16. *New Eng J Med* 326: 1509-1513

Ranum LPW, Duvick LA, Rich SS, Schut LJ, Litt M, Orr HT (1991) Localization of the autosomal dominant HLA-linked spinocerebellar ataxia (SCA1) locus, in two kindreds, within an 8-cM subregion of chromosome 6p. *Am J Hum Genet* 49: 31-41

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

Reeders ST, Keith T, Green P, Germino GG, Barton NJ, Lehmann OJ, Brown VA, Phipps P, Morgan J, Bear JC and Parfrey P (1988) Regional localization of the autosomal dominant polycystic kidney disease locus. *Genomics* 3: 150-155

Richards RI, Holman K, Lane S, Sutherland GR and Callen DF (1991) Human chromosome 16 physical map: mapping of somatic cell hybrids using multiplex PCR deletion analysis of sequence tagged sites. *Genomics* 10: 1047-1052

Richards RI, Holman K, Friend K, Kremer E, Hillen D, Staples A, Brown WT, Goonewardena P, Tarleton J, Schwartz C and Sutherland GR (1992) Evidence of founder chromosomes in fragile X syndrome. *Nature Genetics* 1: 257-260

Richards RI, Holman K, Yu S and Sutherland GR (1993) Fragile X syndrome unstable element, p(CCG)n, and other simple tandem repeat sequences are binding sites for specific nuclear proteins. *Hum Mol Genet* 2: 1429-1435

Rogaev EI and Keryanov SA (1992) Unusual variability of the complex dinucleotide repeat block at the SPN locus. *Hum Mol Genet* 1: 657

Rogers DB, Shohat M, Petersen GM, Bickal J, Congleton J, Schwade AD and Rotter JI (1989) Familial Mediterranean fever in Armenians: autosomal recessive inheritance with high gene frequency. *Am J Med Genet* 34: 168-172

Royle AJ, Armour JA, Crosier M and Jeffreys AJ (1993) Abnormal segregation of alleles in CEPH pedigree DNAs arising from allele loss in lymphoblastoid DNA. *Genomics* 15: 119-122

Sack GH (1988) Serum amyloid A (SAA) gene variations in familial Mediterranean fever. *Mol Biol Med* 5: 61-67

Sack GH, Talvot Jr. CC, McCarthy BG, Harris EL, Kastner D, Gruberg L and Pras M (1991) Exclusion of linkage between familial Mediterranean fever and the human serum amyloid A (SAA) gene cluster. *Hum Genet* 87: 506-508

Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA and Arngeim N (1985) Enzymatic amplification of β−globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* 239: 487-491

Sanger F, Nicklen S and Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* 74: 5463-5467

Schellenberg GD, Bird TD, Wijsman EM, Orr HT, Nemens E, White JA, Bonnycastle L, Weber JL, Alonso ME, Potter H, Heston LL and Martin GM (1992) Genetic linkage evidence for a familial Alzheimer's disease locus on chromosome14. *Science* 258: 668-671

Serikawa T, Kuramoto T, Hilbert P, Mori M, Yamada J, Dubay CJ, Lindpainter K, Ganten D, Guenet J-L, Lathrop GM and Beckmann JS (1992) Rat gene mapping using PCR-analysed microsatellites. *Genetics* 131: 701-721

Shen Y, Holman K, Thompson A, Kozman H, Callen DF, Sutherland GR and Richards RI (1991) Dinucleotide repeat polymorphism at the *D16S288* locus. *Nucl Acid Res* 19: 5445.

Shen Y, Thompson AT, Holman K, Callen DF, Sutherland GR and Richards RI (1992) Four dinucleotide repeat polymorphisms on human chromosome 16 at *D16S289, D16S318, D16S319* and *D16S320. Hum Mol Genet* 1: 773

Shen Y, Holman K, Doggett NA, Callen DF, Sutherland GR and Richards RI (1993a) Five dinucleotide repeat polymorphisms on human chromosome 16q24.2-q24.3. *Hum Mol Genet* 2: 1504

Shen Y, Holman K, Doggett NA, Callen DF, Sutherland GR and Richards RI (1993b) Six dinucleotide repeat polymorphisms on human chromosome 16q12.1-q24.1. *Hum Mol Genet* 2: 1505

Shen Y, Holman K, Doggett NA, Callen DF, Sutherland GR and Richards RI (1993c) Three dinucleotide repeat polymorphisms on human chromosome 16p13.11-p13.3. *Hum Mol Genet* 2: 1506

Shen Y, Holman K, Doggett NA, Callen DF, Sutherland GR and Richards RI (1993d) Four dinucleotide repeat polymorphisms on human chromosome 16. *Hum Mol Genet* 2: 1745

Shen Y, Holman K, Doggett NA, Callen DF, Sutherland GR and Richards RI (1994) Dinucleotide repeat polymorphisms at the *D16S525, D16S359, D16S531* and *D16S522* loci. *Hum Mol Genet* in press

Shen Y, Kozman MH, Thompson A, Phillips HA, Holman K, Nancarrow J, Lane S, Chen L-Z, Apostolou S, Doggett N, Callen DF, Mulley JC, Sutherland GR and Richards RI (1994) A PCR-based genetic linkage map of human chromosome 16. Submitted.

Shohat M, Shohat T, Rotter JI, Schlesinger M, Petersen GM, Pribyl T, Sack G, Schwabe AD and Korenberg JR (1990a) Serum amyloid A and P protein genes in familial Mediterranean fever. *Genomics* 8: 83-89

Shohat T, Shohat M, Petersen GM, Sparkes RS, Langfield D, Bickal J, Korenberg JR, Schwabe AD and Rotter JI (1990b) Genetic marker family studies in familial Mediterranean fever (FMF) in Armenians. *Clin Genet* 38: 332-339

Shohat M, Bu X, Shohat T, Fischel-Ghodsian N, Magal N, Nakamura Y, Schwabe AD, Schlezinger M, Danon Y and Rotter JI (1992) The gene for familial Mediterranean fever in both Armenians and non-Ashkenazi Jews is linked to the α-globin complex on 16p: evidence for locus homogeneity. *Am J Hum Genet* 51:1349-1354

Slightom JL, Blechl AE and Smithies O (1980) Human fetal $^{G}\gamma$- and $^{A}\gamma$-globin genes: complete nucleotide sequences suggest that DNA can be exchanged between these duplicated genes. *Cell* 21: 627-638

Smith SA, Easton DF, Evans DGR and Ponder BAJ (1992) Allele losses in the region 17q12-21 in familial breast and ovarian cancer involve the wild-type chromosome. *Nature Genetics* 2: 128-131

Sohar E, Gafni J, Pras M, Heller H (1967) Familial Mediterranean fever: a survey of 470 cases and review of the literature. *Am J Med* 43: 227-253

Somlo S, Wirth B, Germino GG, Weinstat-Saslow D, Gillespie GAJ, Himmelbauer H, Steevens L,Coucke P, Willems P, Bachner L, Coto E, Lopez-Larrea C, Peral B, Millan JLS, Saris JJ, Breuning MH, Frischauf A-M and Reeders ST (1992) Fine genetic localization of the gene for autosomal dominant polycystic kidney disease (PKD) with respect to physically mapped markers. *Genomics* 13: 152-158

Stallings RL, Tornsy DCS, Hildebrand CE, Longmire JL, Deaven LL, Jett JH, Doggett NA and Moyzis RK (1990) Physical mapping of human chromosomes by repetitive sequence finger printing. *Proc Natl Acad Sci USA* 87: 6218-6222

Stallings RL, Ford AF, Nelson D Torney DC, Hildebrand CE and Moyzis RK (1991) Evolution and distribution of (GT)n repetitive sequences in mammalian genomes. *Genomics* 10: 807-815

Stallings RL, Doggett NA, Callen DF, Apostolou S, Chen LZ, Nancarrow JK, Whitmore SA, Harris P, Michison H, Breuning M, Saris J, Fickett J, Cinkosky M, Torney D, Hildebrand CE and Moyzis RK (1992) Evaluation of a cosmid contig physical map of human chromosome 16. *Genomics* 13:1031-1039

Strand M, Prolla TA, Liskay RM and Petes TD (1993) Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. *Nature* 365: 274-276

Straub RE, Speer MC, Luo Y, Rojas K, Oberhauser J, Ott J and Gillian TC (1993) A microsatellite genetic linkage map of human chromosome 18. *Genomics* 15: 45-56

Sutherland GR (1979) Heritable fragile sites on human chromsomes. II. Distribution, phenotypic effects and cytogenetics. *Am J Hum Genet* 31: 136-148

Sutherland GR and Hecht F (1985) "Fragile Sites on Human Chromosomes." Oxford Univ. Press, New York.

Sutherland GR (1988) The role of nucleotides in human fragile site expression. *Mut Res* 200: 207-213

Sutherland GR (1993) Human Fragile Sites. In "Genetic Maps. Locus Maps of Human Maps" (S.J. O'Brien, Ed), Sixth Edition, pp. 264-267, Cold Spring Harbor Laboratory Press, New York.

Tautz D (1989) Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucl Acid Res* 17: 6463-6471

Thibodeau SN, Bren G and Schaid D (1993) Microsatellite instability in cancer of the proximal colon. *Science* 260: 816-819

Thompson AD, Shen Y, Holman K, Sutherland GR, Callen DF and Richards RI (1992) Isolation and characterisation of (AC)n microsatellite genetic markers from human chromosome 16. *Genomics* 13: 402-408

Tomfohrde J, Wood S, Schertzer M, Wagner MJ, Wells DE, Parrish J, Sadler LA, Blanton SH, Daiger SP, Wang Z, Wilkie PJ and Weber JL (1992) Human chromosome 8 linkage map based on short tandem repeat polymorphisms: effect of genotying errors. *Genomics* 14: 114-152

Trent RJ (1993) Forensic medicine. In "Molecular Medicine. An Introductory Text for Students", First Edition, pp. 177-192, Churchill Livingstone, UK.

Tsui LC, Buchwald M, Barker D, Braman JC, Knowlton R, 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 and Donis-Keller H (1985) Cystic fibrosis locus defined by a genetically linked polymorphic DNA marker. *Science* 230: 1054-1057

Wall WJ, Williamson R, Petrou M, Papaioannou D and Parkin BH (1993) Variation of short tandem repeats within and between populations. *Hum Mol Genet* 2: 1023-1029

Wang Z and Weber JL (1992) Continuous linkage map of human chromosome 14 short tandem repeat polymorphisms. *Genomics* 13: 532-536

Weber JL and May PE (1989) Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am J Hum Genet* 44: 388-396

Weber JL (1990a) Informativeness of human (dC-dA)n.(dG-dT)n polymorphisms. *Genomics* 7: 524-530

Weber JL, Kwited AE and May PE (1990b) Dinucleotide repeat polymorphisms at the *D16S260, D16S261, D16S265, D16S266* and *D16S267* loci. *Nucl Acid Res* 18: 4034

Weber JL, Polymeropoulos MH, May PE, Kwitek AE, Xioa H, McPherson JD and Wasmuth JJ (1991) Mapping of human chromosome 5 microsatellite DNA polymorphisms. *Genomics* 11: 695-700

Weber JL and Wong C (1993) Mutation of human short tandem repeats. *Hum Mol Genet* 2: 1123-1128

Weissenbach J, Gyapay G, Dib CS, Vignal A, Morissette J, Millasseau P, Vaysseix G and Lathrop M (1992) A second-generation linkage map of the human genome. *Nature* 359: 794-801

Weissenbach J (1993) Microsatellite polymorphisms and the genetic linkage map of the human genome. *Current Opinion in Genetics and Development* 3: 414-417

Wevrick R and Willard HF (1989) Long-range organization of tandem arrays of α satellite DNA at the centromeres of human chromosomes: high-frequency array-length polymorphism and meiotic stability. *Proc Natl Acad Sci USA* 86: 9394-9398

White RL and Laulouel JM (1988) Sets of linked genetic markers for human chromosome. *Ann Rev Genet* 22: 259-279.

Whitmore SA, Apostolou S, Lane S, Nancarrow JK, Phillips HA, Richards RI, Sutherland GR and Callen DF (1993) Isolation and charaterisation of transcribed sequences from a chromosome 16 hn-cDNA library and the physical mapping of genes and transcribed sequences using a high resolution somatic cell panel of human chromosome 16. *Genomics* in press

Wilkie AOM, Higgs DR, Rack KA, Buckle VJ, Spurr NK, Fischel-Ghodsian N, Ceccerini I, Brown WRA and Harris PC (1991) Stable length polymorphism of up to 260 kb at the tip of the short arm of human chromosome 16. *Cell* 64: 595-606

Wilkei PJ, Krizman DB and Weber JL (1992) Linkage map of human chromosome 9 microsatellite polymorphisms. *Genomics* 12: 607-609

Wilkie PJ, Polymeropoulos MH, Trent JM, Small KW and Weber JL (1992) Genetic and physical map of 11 short tandem repeat polymorphisms on human chromosome 6. *Genomics* 15: 225-227

Willard HF, Waye JS, Skolnicsk MH, Schwatz CE, Powers VE and England SB (1986) *Proc Natl Acad Sci USA* 83: 5611-5615

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

Yan W, Boustany R-MN, Konradi C, Ozelius L, Lerner T, Trofatter JA, Julier C, Breakefield XO, Gusella JF and Haines JL (1993) Localization of Juvenile, but not late-infantile, Neuronal Ceroid Lipofuscinosis on chromosome 16. *Am J Hum Genet* 52: 89-95

Yule GU and Kendall MG (1968) "An introduction to the theory of statistics," 14th ed., Charles Griffin, London.

Zoghbi HY, Jodice C, Sandkuijl LA, Kwiatkowski TJ Jr, McCall AE, Huntoon SA, Lulli P, Spadaro M, Litt M, Cann HM, Frontali M and Terrenato L (1991) The gene for autosomal dominant spinocerebellar ataxia (SCA1) maps telomeric to the HLA complex and closely linked to the D6S89 locus in three large kindreds. *Am J Hum Genet* 49: 23-30

Zuliani G and Hobbs HH (1990) A high frequency of length polymorphisms in repeated sequences adjacent to Alu sequences. *Am J Hum Genet* 46: 963-969

# APPENDIX I

## (AC)n·(TG)n REPEAT SEQUENCES

*D16S525* (AC308G7)

5'GCTGCGGGCGAGACCTCAGGCGCGAGACGGGCTGTGTGTGTGTGTGTGTGTGTGAGAGA
GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGTGTGTGTGTGTGTGTGTGTGTGT
GTGTGTGTGTGTGTGTGTCGGGGGAGGGGNGGCTCCTGGGCGAAGGATCGACTCGCAGTC
CTGCCCCCC 3'

*D16S523* (AC13H1)

5'CCCCACAGTGTTCCATTTTCTCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTAGN
GTGTGTGACAAAGTCTNTCTCTGCTACCCAGGCA 3'

*D16S453* (AC301G12)

5'TAAAAAGCTGTGGACTCATGGGTCAGGACTTAATAAAGTTGATTTTTACTGCTTTTTCTG
ATGCAACACTGGTGTGTGTGTGTGTGTGTGTGTATCGTGTGTGTGGTGGTGAAGACTCAAT
TACAACAAGACCAGCACAGTGTGGTGCCAGTCTTGAGTCAGTGCAGG 3'

*D16S454* (AC45G5)

5'CAAAGGCTTGAGCTGCCATCCTAAGCTTTGGTTAGGAGTATGTGTGTGTGTGTGTGTG
TGTGTGTGTGTGTGTGTGTGTGTGTATGTGTATGTGTCTGTGTGTGTTGGATTTGCAGGCTN
ACTCCCGGTATTTAGNCTCATAGATATCTAGNNTTCATTTTTGGCAAGTTGAAGATCCTCT
AGAGTCGACCTGCAGGAATTNCAAGCTTGGAGCA 3'

*D16S49* (ACCRI.0114)

5'AATTCAAGGGAGGCTCATGTGAANNAGGTCAATGGGTGGGTGTGTGTGTGGGTGGGT
GTGTGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTTCAAATCCCAGTTCCTCATAACATAGA
GGGAGGAGTGTTGG 3'

The underlined sequences are primer sequences.

(AC7.22)

5'GGGATTACAGTCGTGAGCCACCATGCCCAGCGTAGGTAATTTTTAAGATATTCATTAATT
TCAATTTTGTCATTTTCTCTTTTGCATTTTTAAGCCAATGTGTGTGTGTGTATGTGTGTTTGT
GTGGTGTTTTTTGTTTGTTTGTTTGTTTGTTTGTTTGTTTTTGGTCTCAAACATTTCTGTAGAT
C 3'

*D16S524* (AC40A7)

5'CCAGTCTCCAGACTATGGGGGAGAAGGCTAAAGATTCATTTGTGAGTAAAAGTGTCTTT
AGAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGGNCTTAAAGATATCTGCTT
GAAAGGTTGAATTCTNTCCTTTGAAACTCAAAGGAATAGAGTAACTGTTACCTGTTGCACA
CCTAAACATTCAAAACCA 3'

*D16S452* (AC33A4)

5'CAAGAGAACCASAATGAGTTGCTTCTTGAACCTCCCACTTTACCCTTAGGGCTCTCCTCT
TCCCAGGTGTGTGTGTGTGTGTGTGTGGTGTGTGTGTGTGTGTGTGTGTGTATGTGTGTG
TGTATGTGTGTACAGGGGAACAACACCATGGNCACTGA 3'

*D16S370* (AC25H11)

5'TCTTAGATGTTTCAGGGCAACTAGATATGCCCATACGATTGCCCATATTCTGTGTGTGTG
TGTGTGCGTGTGTGTGTGTGTGTGTTTAATTGGAGCTAGGCATTAGTTACTTAGNCAATGAG
T 3'

*D16S319* (AC7.14)

5'ATCGCGCCCAACCCACTATTGTTTTTTTTTTGTGTTTTTTGGGTTTGTGTGTGTGTGTGTGT
GTGTGTGTGTGTGTGTGTGGTTTTTTTTGTTTTTTTTTGTTTTGTTTTGTTTTGTTTTTTTTGTGG
GGAGGGGGACGGAGTTTCACTCTTGTCACCCAGGCT 3'

*D16S288* (AC7.1)

5'GGATTTCTTTGCTTTCAAATATTTCTTTTGTAACCAAAAGTGAAAGTCTGGCAAATTCCC
CTAAATCACAGACTCCCTTGGCTCTTCTGCCCCCACCCTGCCCCTCCCTCCCTATTTTCCA
GTGAGTTTTCAAAACCAACAGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT
GTGTACATCTCTGGGGGAAGAGACACCTCTGCTCCTCATCCCTCTCAGAGTTGCCCCATC
CATCAGCCTGGGTTTGC 3'

*D16S383* (AC80B3)

5'ACTCCTCTCCCCATAGAAGGCATCAAACCTGCAGGAAGAAAGAAAGAAATTAGCGTGT
GTGTGTGTGTGTGTGTGTGTGTGTGTAAGGGGTGAGTGGGTGGGGGTGGTAAGGGATCTGC
TCTCGTCTGACCTTGATTTAAGAGAACGTCTCTATCC 3'

*D16S359* (AC26E3B)

5'GGGAGCTTGCTTGGGGTTAAATGCGTGTGTGCACCCCGGCAGTGTGTGTGTGTGTGTGTG
TGTGTGCGCGCGCGTGCATGCAACGCACATGCATACACATGCATGCATGC 3'

*D16S390* (AC10F5)

5'GGACAGCCTCAAATGAAATCTCACACCTTGGTTAGCAGTGGAAGGTTGCCGCGCTCAAT
ATTATTTGTCTTAGCATGTGTGTGTGTGTGTGTGTGTGTGTGTGCGCGCGTGTGTGTGTG
TGTTTCTTCTCCCTCCTCTTAAGATCTTATCAAATCATAGTTATAAACAACTACCCTAGCT
GAGAGTCTATATTC 3'

*D16S531* (AC8.15)

5'GATCCTCTAGAGTCGTTCTCACACACCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG
TGTGTGTGTGTGTGTGTGTGTGTGTGTAGGGGGTGTGCCTTTTGCTTACTGTGTCTGCTA
TCTGGCTGAGATTCAGATATCACGCTCCCCACT 3'

*D16S317* (AC7.9)

5'CTCCAGACAGTGCTTGGATAAAGTGGGACATAACTACATGAGTCAGGAGTACACAGATG
CACCAGACCTGTCCCTCTTCCAGAGCATCACTAAAGGAGGCAGACGTGTGTGTCCCTGTG
TGTGTTATGTGTGTGTGTGTGTGTGTGTGTGTCAGGGTGGGACCTACTAGATTTCTTATG
ACTGCAAAGCCAGGGTATCAAAGGGAGATC 3'

*D16S320* (AC8.52)

5'AAAAAAAAAAAAAAAAAGTCTGAGAGACATCCAGGTTTGAAATGTGCGTGTGTGTGTGTG
TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGAGACTTAGGGAGAGGTCTACAGGAG
ATAGAAATTTTGGAGGCCTCAGCTTGCAGATGATGCTTAAAGCCACAGGACTGACTGATA
TCAC 3'

*D16S389* (AC10B3)

5'CCTCCAAATGTGGTCCTGATATTTTAATTTAGCAATCCAGTGAGATACTCGTGTGTGTGT
GTGTGTGTGTGTGTGTGTGTGTCTCCATAACATTCCTGATGGAAATTAG 3'

*D16S451* (AC69F12)

5'ATATCCATCACCATTCATGGTTACCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG
TGTGTGTGTGTGGTAAAAAAAAATTAAGGTCCTACCTTTCCTTAGT 3'

*D16S347* (AC12F8)

5'ACTAAAAGCAAATACAGAAGCAGCAACATTACATGGTCCAGGATATATATNAANATGTG
GTAGGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTATCTCAGAGGTAGTAGATAG
ATTGTCCTTTTATAGAGATGAGCAAACATTTCTGTAAAGGGCCAGATAGTGGATATGCTCA
GCTAGCACTTATGCGTGGG 3'

D16S318 (AC8.20)

5'TATATATTTTCTGTGGTGTCACATCAGGAAGTACATAATGATTTGGTGAGGTGTGTGTGT

GTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTATGTGGTGTTTATGTATAAGATTGGTCA

ATGGGTTCAGTCCAATCGGTTCATGTGTAGTCTGTTTTTTAAAA 3'

D16S522 (AC8.21)

5'GCCTCAAAAATAAAATAAATTAAAATGAGAGTGAGAATTCCTAGCTTTGGAATTGATGT

GTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTTGGGCGAGGGGTGGTA

GCCATTGTAGCGTGGG 3'

D16S450 (AC80H3)

5'ACATACAAAGTCTCTGGAGGAGTGGTGTTCTCTGCGTGTGTGTGTGTGTGTGTGTGTG

TGTGTGTGTGTGTGTGTAGAGTTGTAGTTATGAGGATGTGGGGGGCCC 3'

D16S395 (AC33G11)

5'ACATTTGCATTCTAACGCAAAATTGAACTCTATGTGTATCTGTCTTTGTATTTGTGTGTGT

GTGTGTGTGTGTGTGTGTGTGTGTTCACTCAGGCCATAACAGGCATGTGCAAGGCA 3'

D16S289 (AC7.46)

5'GATCACCACTTATCATTTCTTCCAAGCTGTGTTTCAAATGTATGGTGTGTGTGTGTGTGTG

TGTGTGTGTGTGTGTGTGTGTGTGTCTAGAGACAAATGCCAGTTTTGTCAGTGCTGAACCA

TACAATTTACATGGCTGAATACCCTGCTTCTCTTCCTCCAACTGCCTTACCTAATAATAAT

TTTCATGG 3'

D16S40 (ACCRI.015)

5'TAGGGCTCTTGATTGNTTGCAGTCTTCATGCCTAACCACACCTCTGAGTGTCTCTGTTGA

CTAGAGCACAGGTATATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTATACGCACCCAGCG

CCCTTTGAGGACC 3'

*D16S363* (AC51G1)

5'CACATGTTGATTCTCATTTACCTGTTTATTAAAAANAAATTTAGNTNCAGGCTCATNTAC
TAATTTCNTCCAGNGCTTAAGGTNTGTNTGNNATTGGGGATAGANTTCAGNGAACAATGNT
ANGGTTTACAAGGGCCCATNTTTGACTTCAGGGCCGCGTGTGTGTGTGTGTGTGTGTGTGT
GTGTGTGTGTGTGTGTGTATACATGTGTATGTGTACATACANTTCCCCCTCTCACAGGATTT
CTTTAGATCCTC 3'

*D16S392* (AC305E9)

5'AGAGATCTTCGGAATACTACTCGTTGAGTGTGTGTGTGTGTGTGTGTGTGTGTATGTAT
ATATATATATATTTTTTTTTTGAGACAGAGTTTCGCTCTTGTCACCCAGGCTGAGTGCAATG
G 3'

*D16S332* (AC305D6)

5'TGAGTCATGTTCATGCCACTACACTCAGCTGGTGACAGACGAAGACCCTGTCTCAAAAA
AATATATACATATGTGTGTGTGTGTGTGTGTGTGTGTGTCAGGAATATATTTCTGTGCCTTT
AAATCAGACCT 3'

D16S393 (AC323H4)

5'CCCTACGCTTCTTCAGTTAGAAGGAAGAGGACATTTATTCATAGGAAGGAGGGAATGAG
AGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTCGTGAAT
TNNGNNANTTAATTTTTTCATATGATTAGNGCCGAAGTCTGGGATCCTCTAGAGTCGCCCT
GNAGGCATGCAAGTTTGGCGTAATCATGGACATA 3'

*D16S449* (AC51A4)

5'GCCTGGGATTCCAGACTGCTGGCCTGCCCTGTGGATTTCAGACTTACCAACCTCCACAG
CCATGTAAACCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTCTGTGT
GCGCGTGGTGGNGGGTGTNCTGNCNGNGTGTNTATGCGTGTGTNTGTATGCGTNTGTGTGT

ATGNGTGTGTGTTCTNTTACTCTGGAAACCCAGCATTGCTAATNCCGCTCGT 3'

(AC7.59)

5'TTACTGAATATTTCTGTGCCTAGGTTTCTTCATGTGAAATGGGGATTGTTGTGAGAACAC

AAAGGGATTCCCAGGGCAGTTCCTAGTGCATAGTCTGGCTGCCTTTGTATGTGTGTGTGTG

TGTGTGTGTGCGTGTGTGTGTGTGTTTAATATAGAGACAGGGTCTCACTCTGTTGCCTAG 3'

The primers of AC7.59 detected multiple loci on human chromosomes 12, 14, 16, 18 and Y.

(AC39C8)

5'ACACAAATTCCTAAAGCGTTTTTCGTGTGTGTGTGTGTGTGTGTGTGTGCGTGCGTGT

GCGTGTGTGTGTGTAGAGATGGGGTTTCGNCATATTGCCCAGGCTGGTCTCGAACTCCTGG

GCTCAAGCGATCCTCTAGAGTCG 3'

The primers of AC39C8 detected multiple loci on human chromosomes 12, 14 and 16.

# APPENDIX II

## PEDIGREES AND GENOTYPES OF

## FMF FAMILIES

This Appendix II lists the genotypes at four (AC)n repeat loci on 16p13.3 in the 62 FMF families. The analysis of these data is presented in Chapter 6.

The four (AC) repeat loci listed in order from the telomere to centromere are:

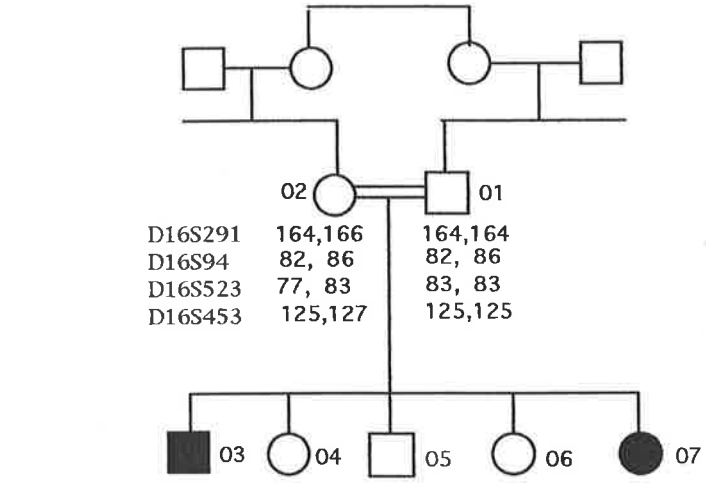
| | |
|---|---|
| D16S291 | AC2.5 |
| D16S94 | ACVK5 |
| D16S523 | AC13H1 |
| D16S453 | AC301G12 |

The pedigrees 1 to 49 are non-Ashkenazi Jews provided by Dr. D. L. Kastner (National Institute of Arthritis and Musculoskeletal and Skin Diseases, Bethesda, USA).

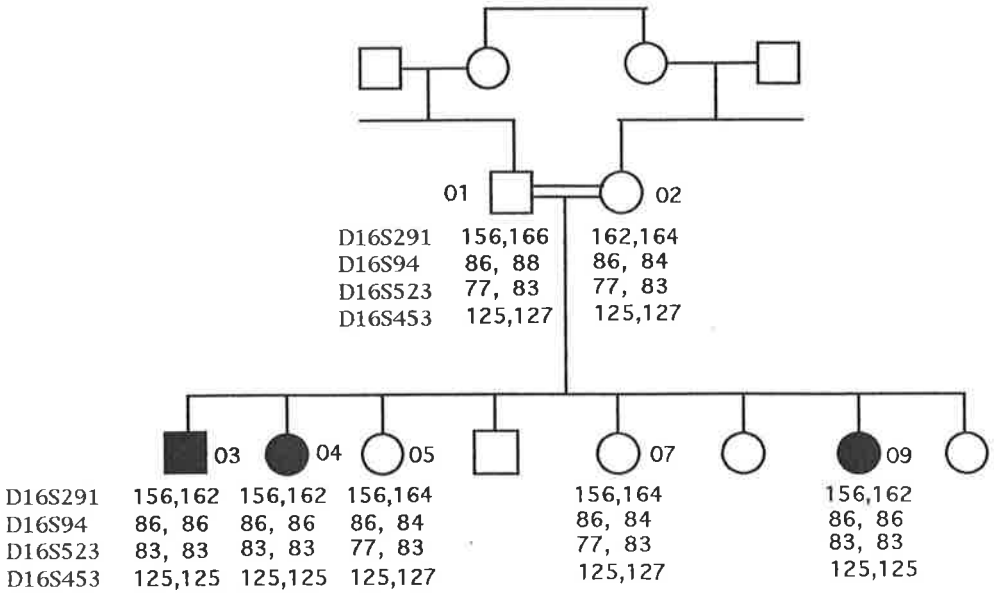
The pedigree 50 to 62 are Armenian provided by Dr. N. Fischel-Ghodsian (Department of Paediatrics of Cedars-Sinai Medical Centre, Los Angeles, USA).

The origin numbers of FMF families are shown in brackets.

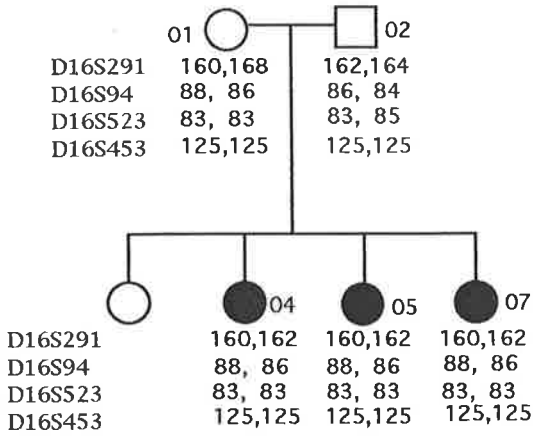
## Pedigree 1 (# 2)



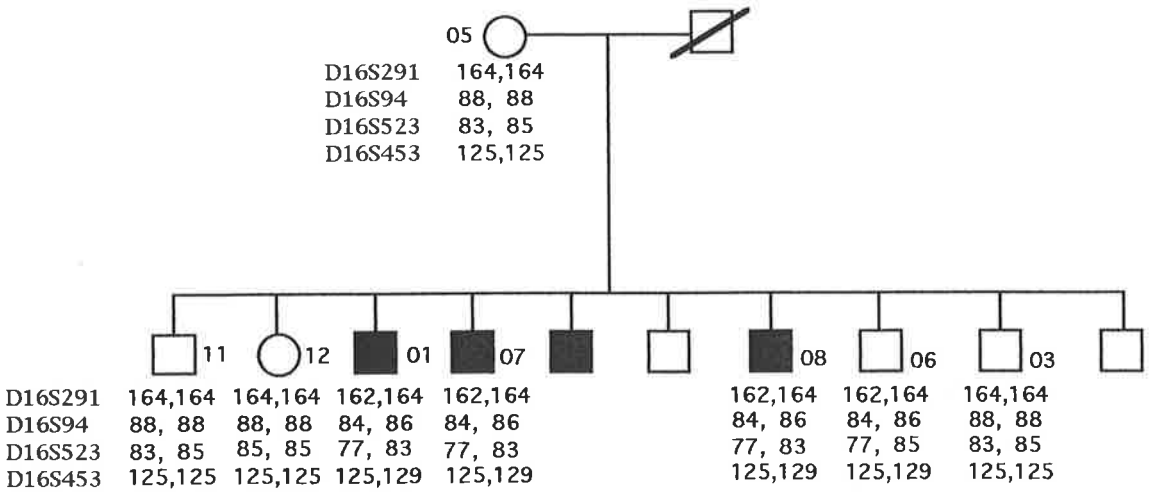

## Pedigree 2 (# 3)



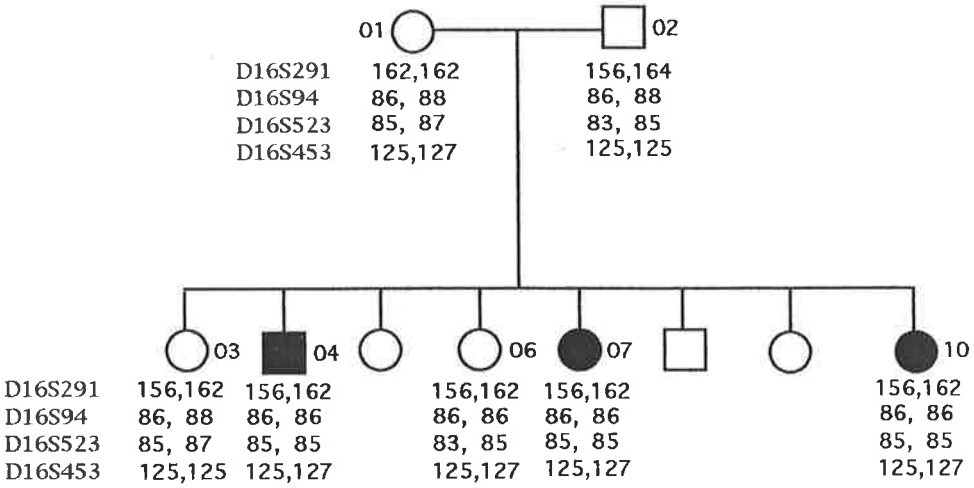
## Pedigree 3 (# 4)



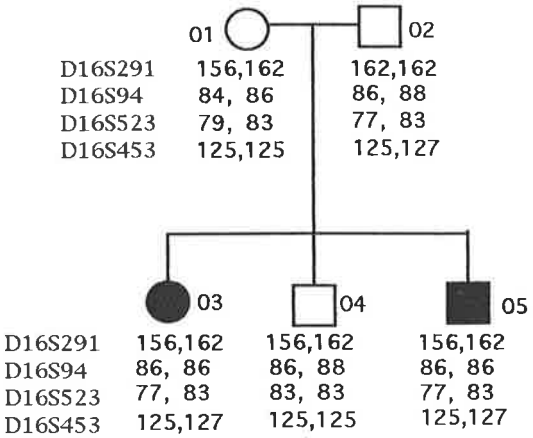

## Pedigree 4 (# 6)



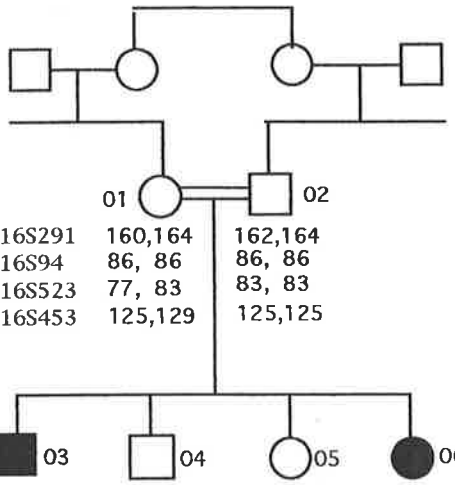
## Pedigree 5 (# 7)



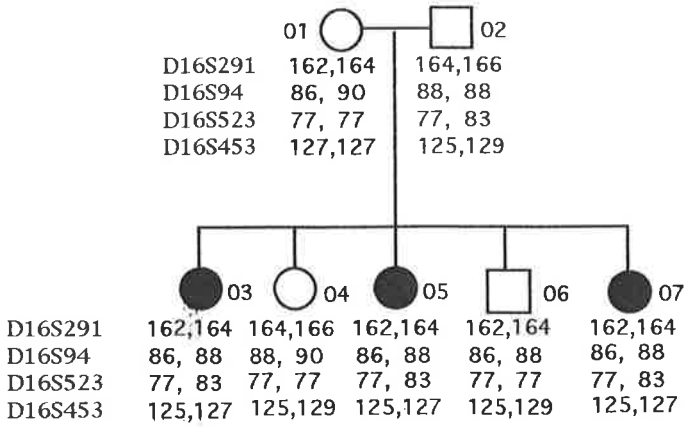

## Pedigree 6 (# 8)



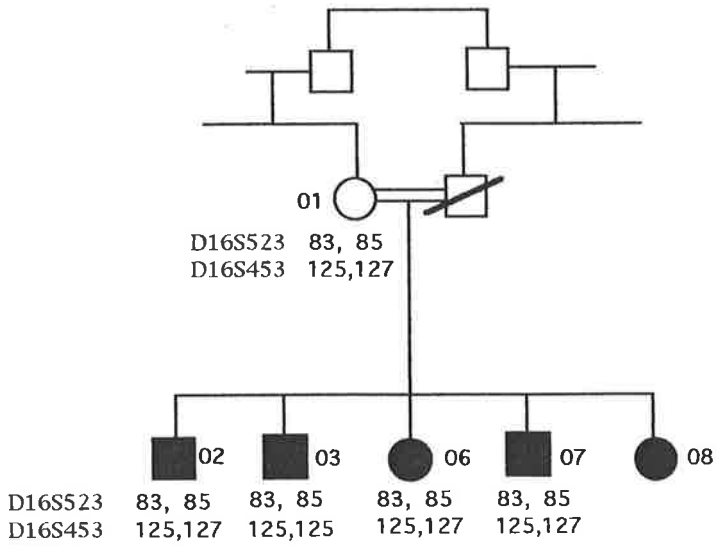
## Pedigree 7 (#10)



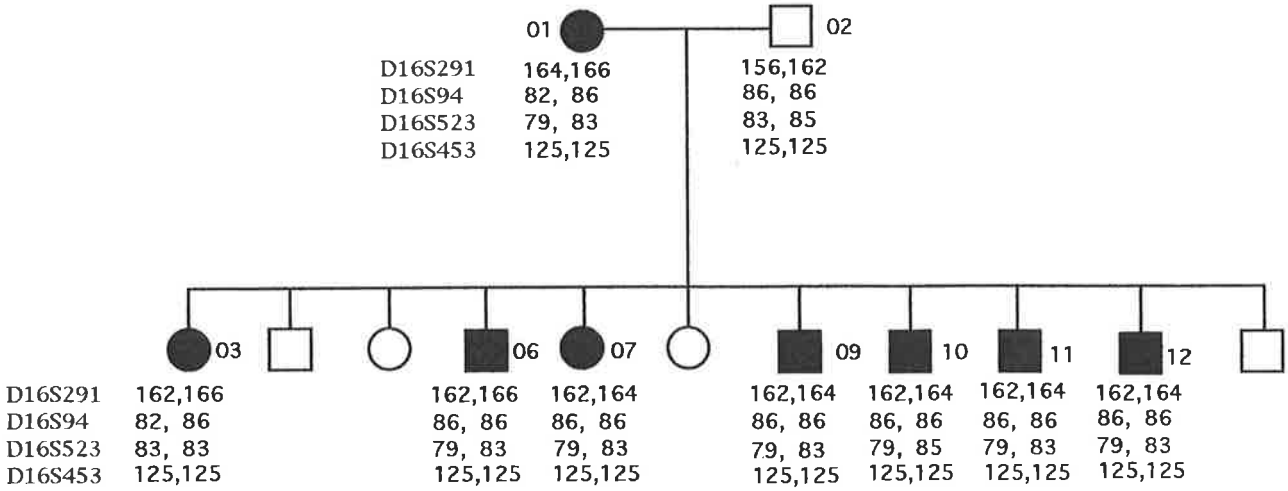

## Pedigree 8 (#11)



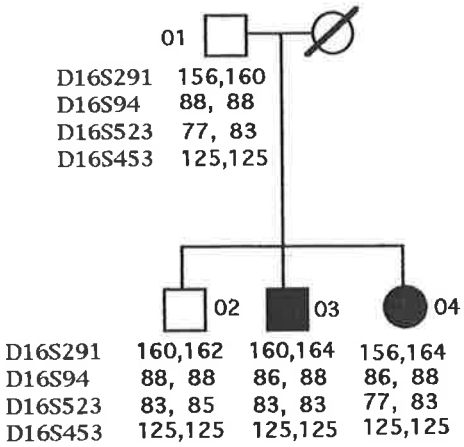
## Pedigree 9 (#12)



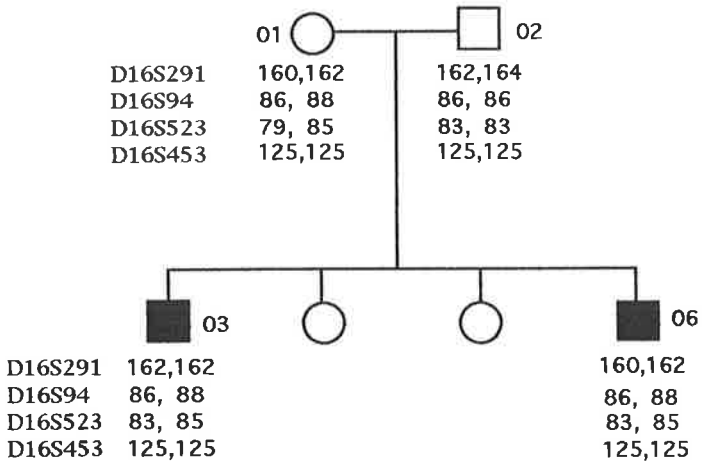

## Pedigree 10 (#13)



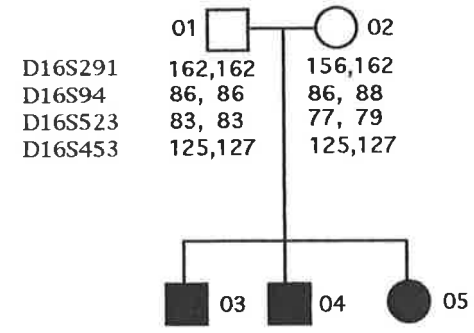
## Pedigree 11 (#14)




## Pedigree 12 (#15)

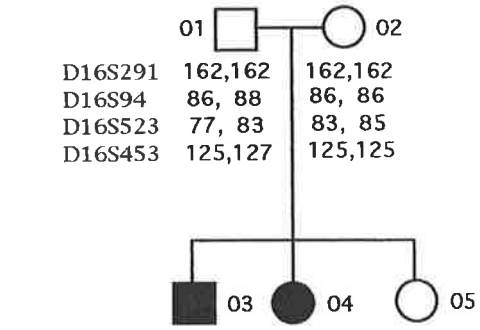


## Pedigreee 13 (#16)



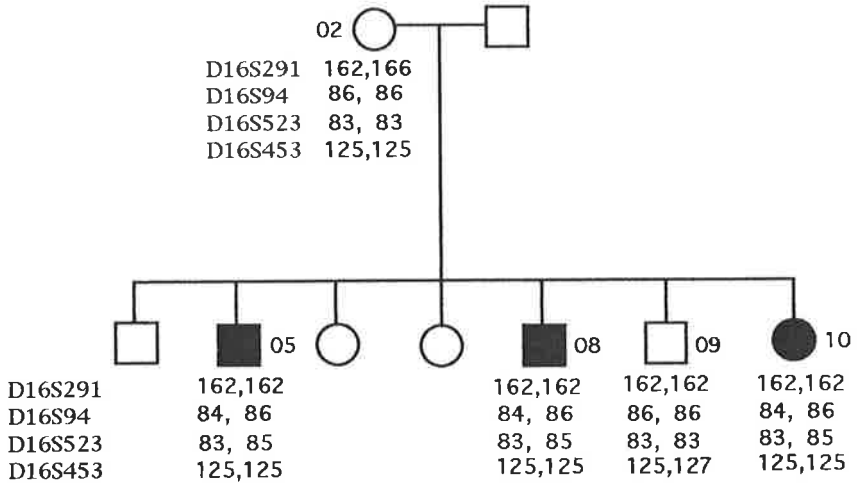

| | 01 | 02 | 03 | 04 | 05 |
|---|---|---|---|---|---|
| D16S291 | 162,162 | 156,162 | 162,162 | 162,162 | 162,162 |
| D16S94 | 86, 86 | 86, 88 | 86, 88 | 86, 88 | 86, 88 |
| D16S523 | 83, 83 | 77, 79 | 79, 83 | 79, 83 | 79, 83 |
| D16S453 | 125,127 | 125,127 | 125,125 | 125,125 | 125,125 |

## Pedigree 14 (#17)

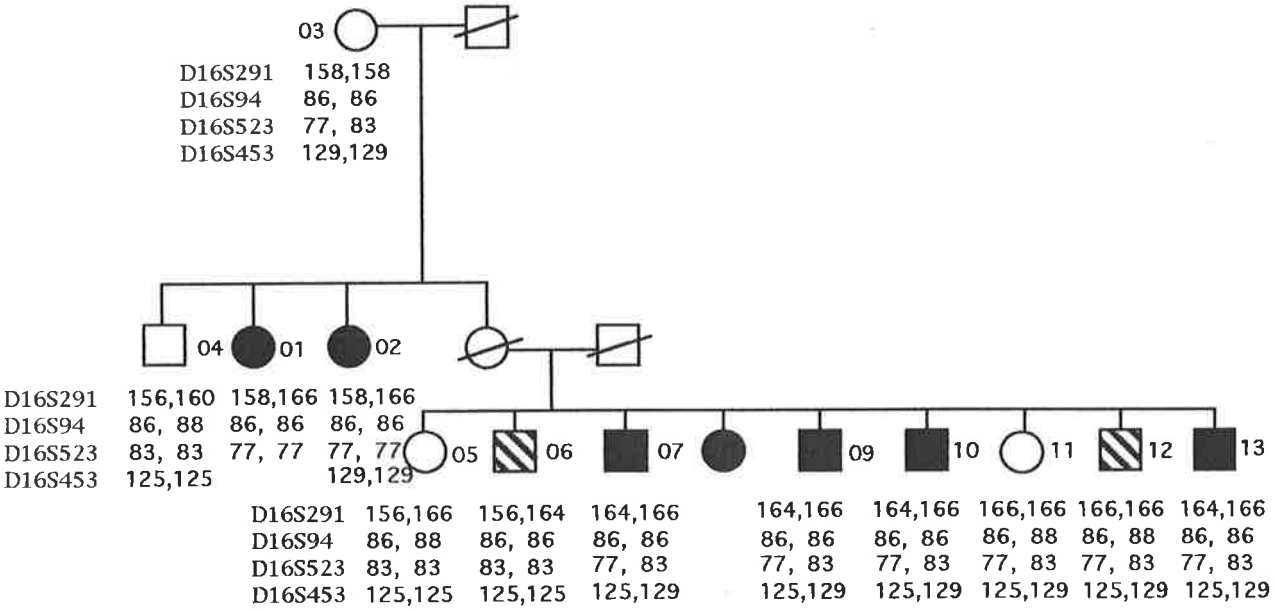



| | 01 | 02 | 03 | 04 | 05 |
|---|---|---|---|---|---|
| D16S291 | 162,162 | 162,162 | 162,162 | 162,162 | 162,162 |
| D16S94 | 86, 88 | 86, 86 | 86, 86 | 86, 86 | 86, 88 |
| D16S523 | 77, 83 | 83, 85 | 83, 83 | 83, 83 | 77, 85 |
| D16S453 | 125,127 | 125,125 | 125,125 | 125,125 | 125,127 |

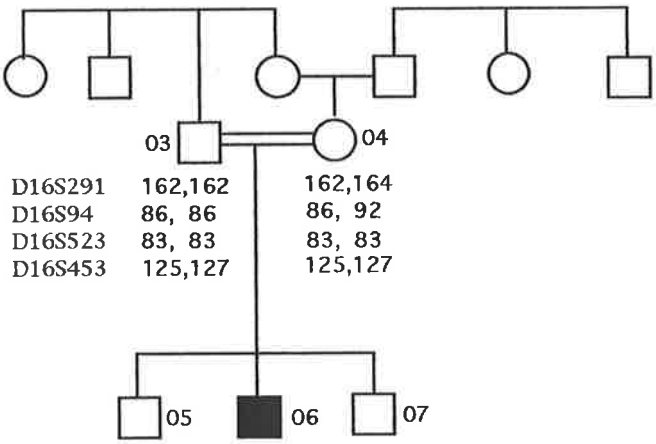
## Pedigree 15 (#21)



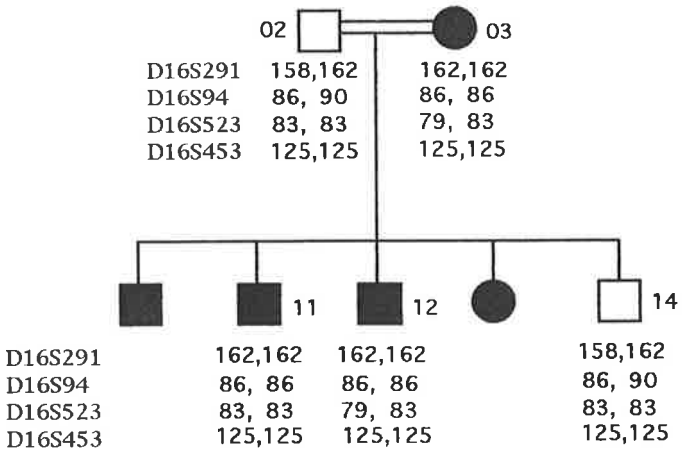

## Pedigree 16 (# 23)



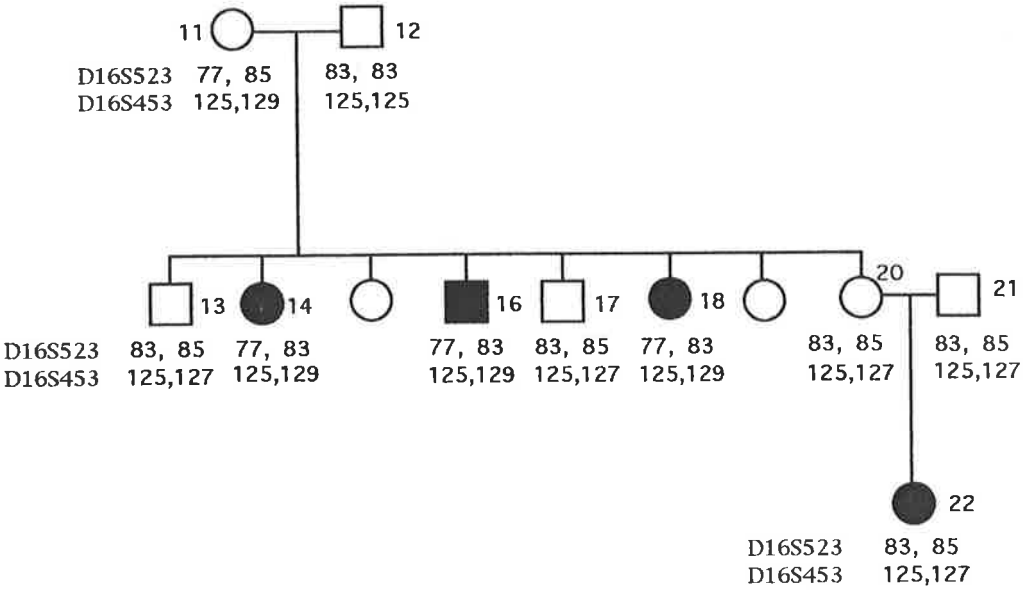
## Pedigree 17 (#25)



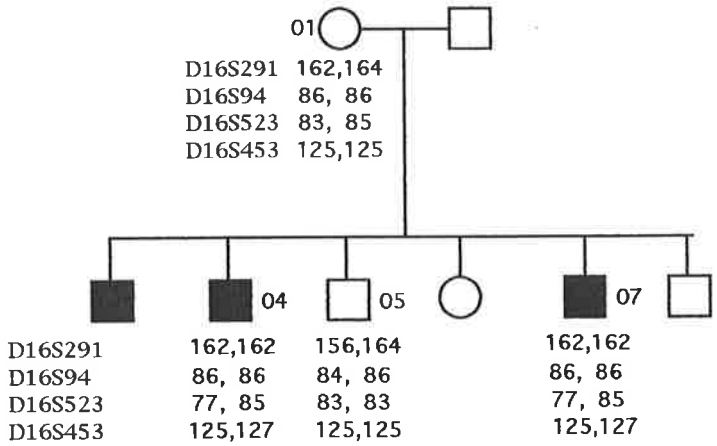

## Pedigree18 (#27)



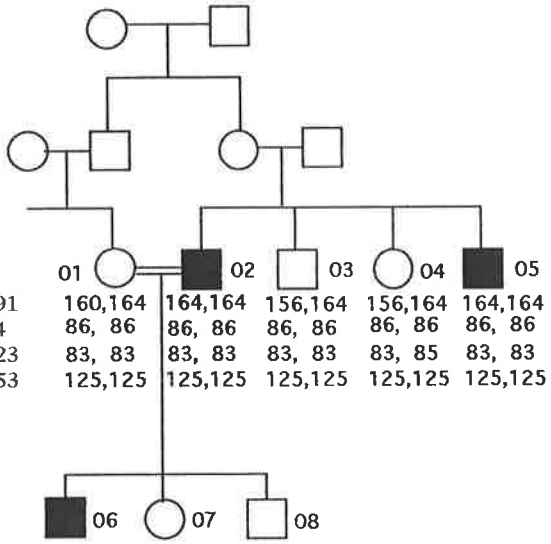
## Pedigree 19 (#29)



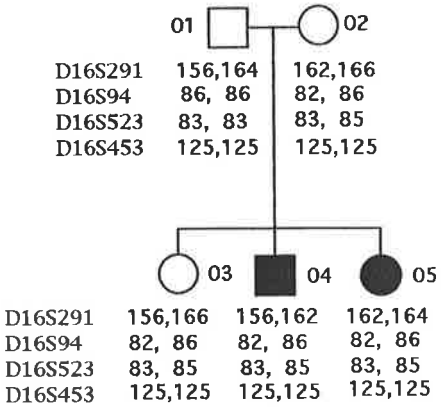

## Pedigree 20 (# 30)



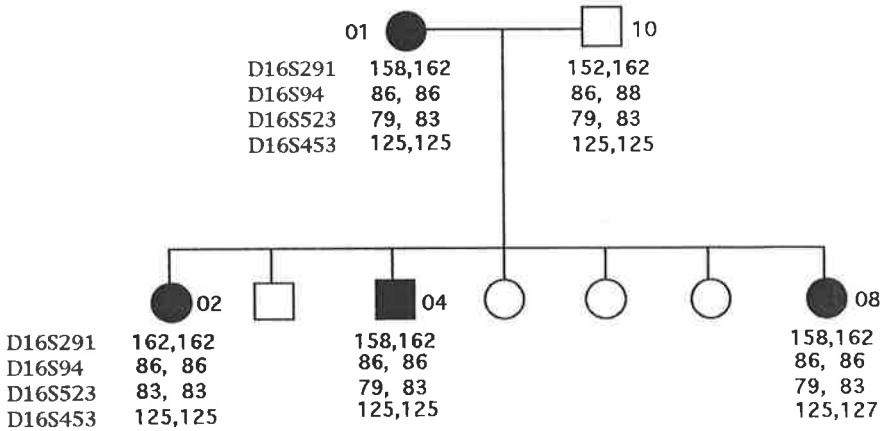
## Pedigree 21 (# 31)



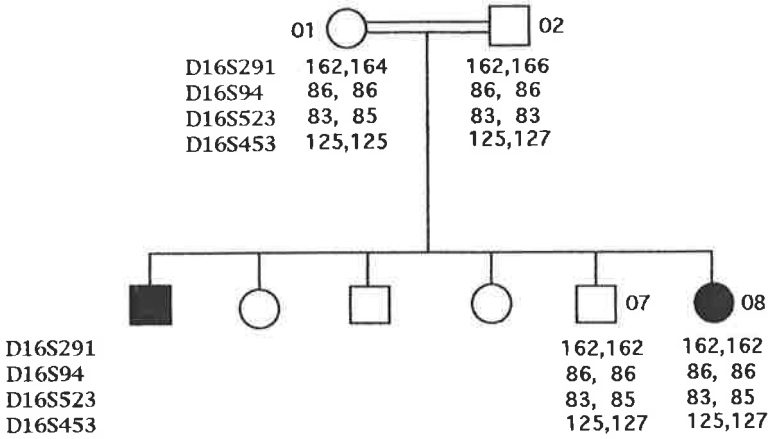
## Pedigree 22 (# 33)



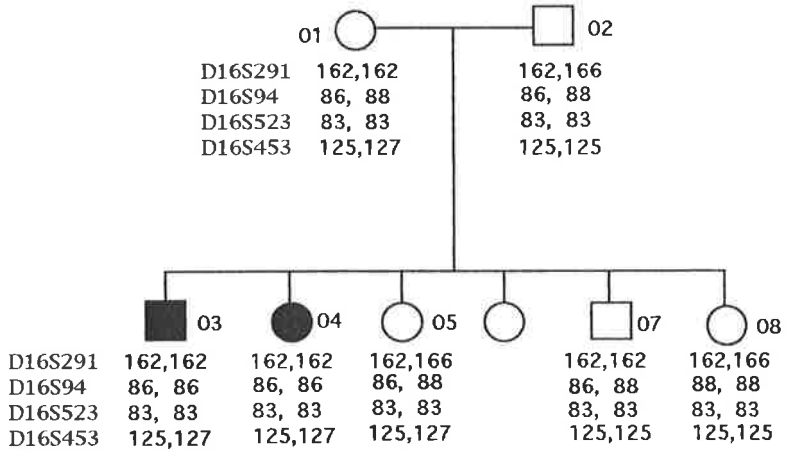

## Pedigree 23 (# 35)



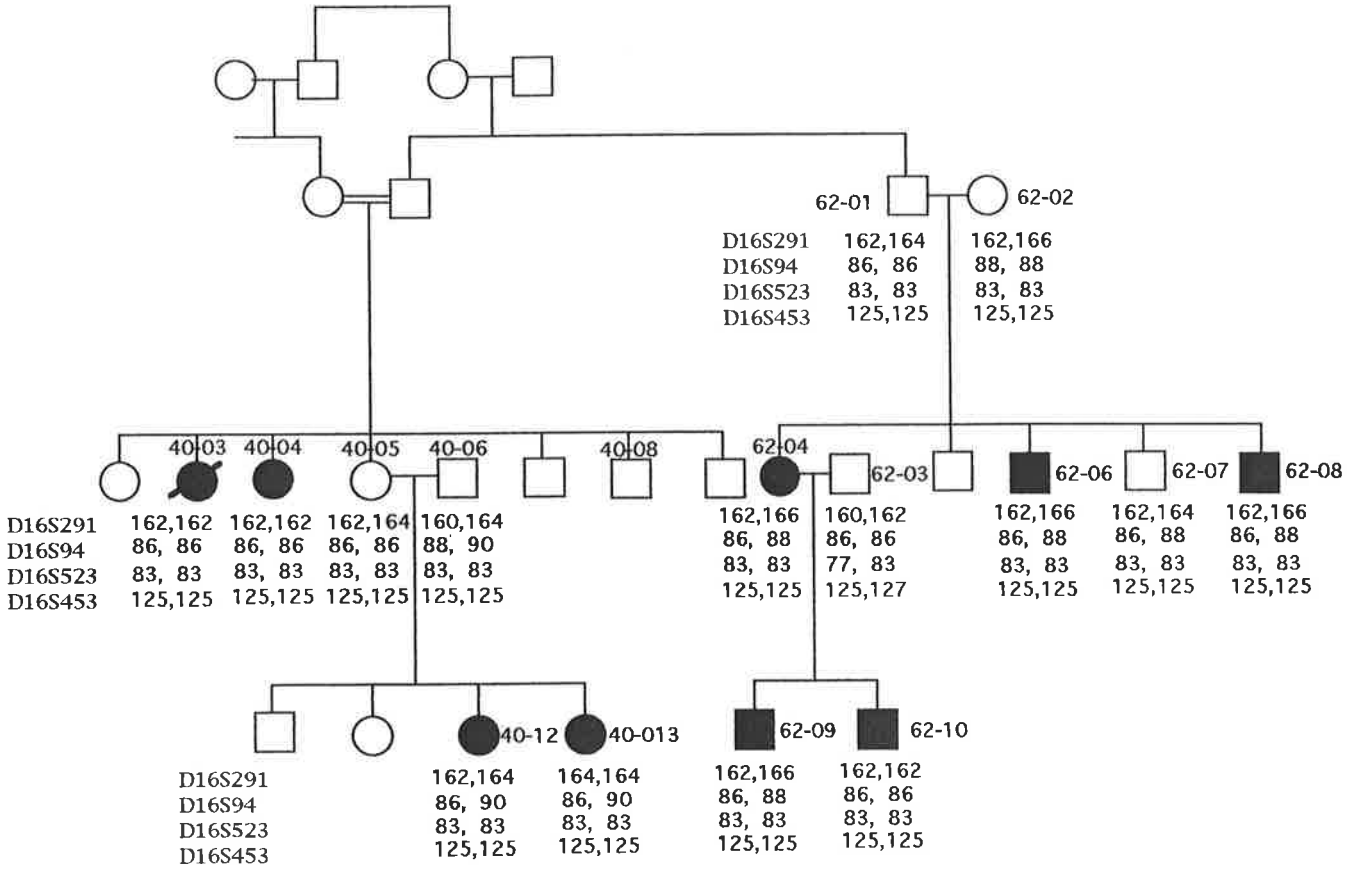
## Pedigree 24 (# 38)



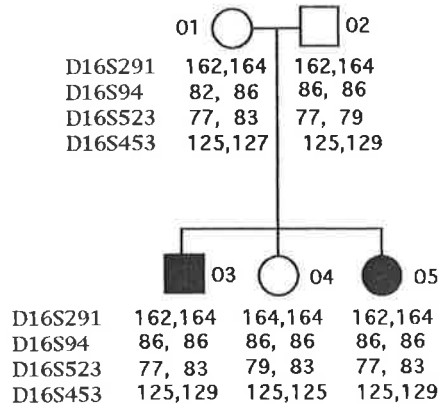

## Pedigree 25 (# 39)



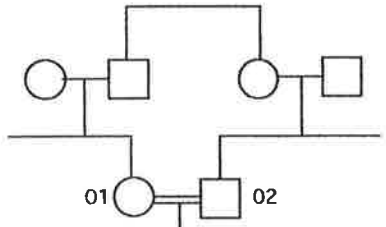
## Pedigree 26 (# 40/62)




## Pedigree 27 (# 42)



## Pedigree 28 (# 46)

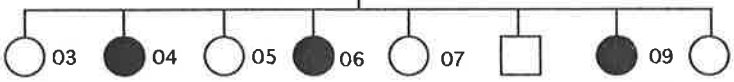



| | 01 | 02 |
|---|---|---|
| D16S291 | 164,164 | 158,164 |
| D16S94 | 86, 90 | 86, 86 |
| D16S523 | 83, 83 | 83, 83 |
| D16S453 | 125,125 | 125,125 |


| | 03 | 04 | 05 | 06 | 07 | 08 | 09 |
|---|---|---|---|---|---|---|---|
| D16S291 | 164,164 | 158,164 | 164,164 | 164,164 | 164,164 | 164,164 | 164,164 |
| D16S94 | 86, 90 | 86, 90 | 86, 90 | 86, 90 | 86, 90 | 86, 90 | 86, 90 |
| D16S523 | 83, 83 | 83, 83 | 83, 83 | 83, 83 | 83, 83 | 83, 83 | 83, 83 |
| D16S453 | 125,125 | 125,125 | 125,125 | 125,125 | 125,125 | 125,125 | 125,125 |

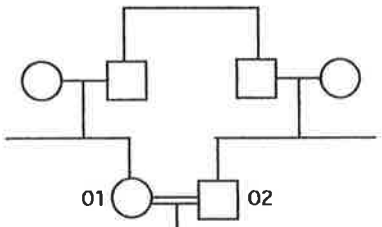
## Pedigree 29 (# 49)


| | 01 | 02 |
|---|---|---|
| D16S291 | 162,164 | 156,162 |
| D16S94 | 86, 86 | 86, 86 |
| D16S523 | 83, 85 | 79, 83 |
| D16S453 | 125,127 | 125,127 |




| | 03 | 04 | 05 | 06 | 07 | 09 |
|---|---|---|---|---|---|---|
| D16S291 | 162,164 | 162,162 | 156,162 | 162,162 | 156,164 | 162,162 |
| D16S94 | 86, 86 | 86, 86 | 86, 86 | 86, 86 | 86, 86 | 86, 86 |
| D16S523 | 79, 85 | 79, 83 | 83, 83 | 79, 83 | 83, 85 | 79, 83 |
| D16S453 | 127,127 | 125,127 | 125,125 | 125,127 | 125,127 | 125,127 |

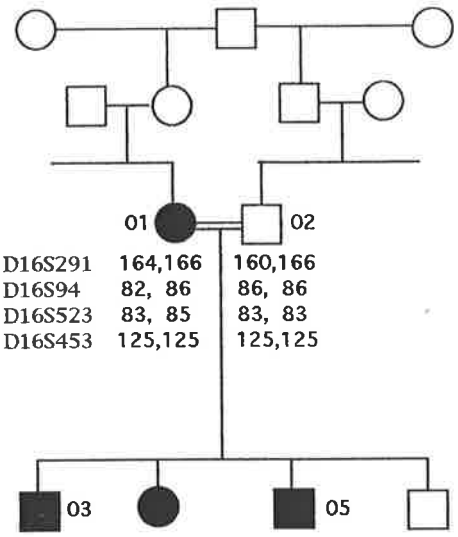
## Pedigree 30 (# 60)



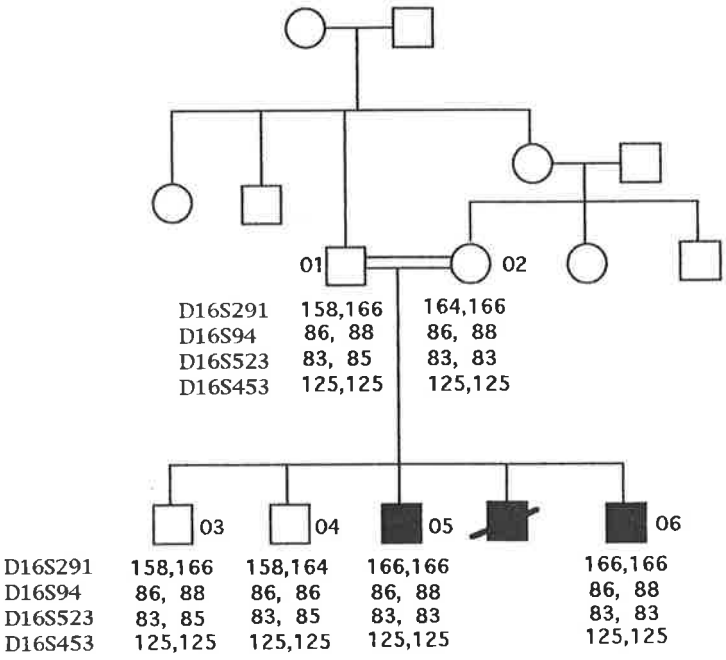

| | | |
|---|---|---|
| D16S291 | 164,166 | 160,166 |
| D16S94 | 86, 86 | 86, 88 |
| D16S523 | 77, 83 | 77, 85 |
| D16S453 | 125,129 | 127,129 |


| | | | | |
|---|---|---|---|---|
| D16S291 | 166,166 | 160,164 | 164,166 | 166,166 |
| D16S94 | 86, 86 | 86, 88 | 86, 86 | 86, 86 |
| D16S523 | — | 83, 85 | 77, 83 | 77, 77 |
| D16S453 | 129,129 | 125,127 | 125,129 | 129,129 |

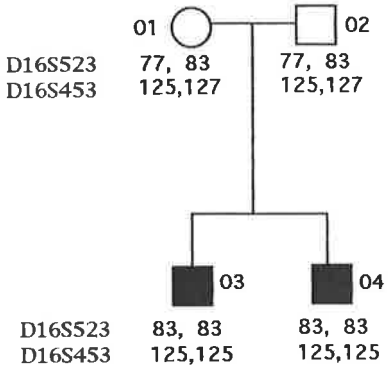
## Pedigree 31 (# 72)



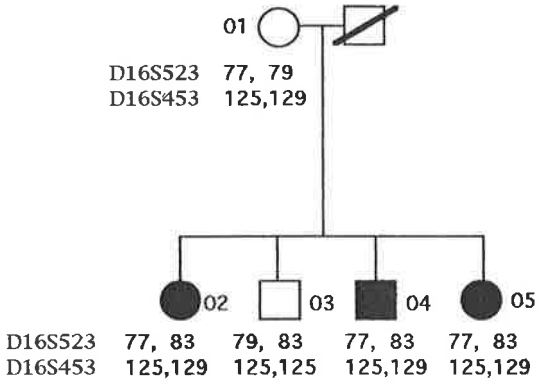

## Pedigree 32 (# 75)



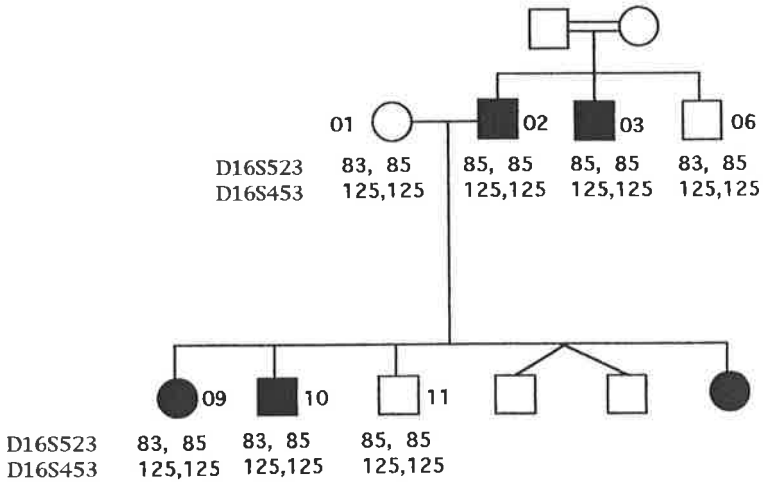
## Pedigree 33 (# 101)



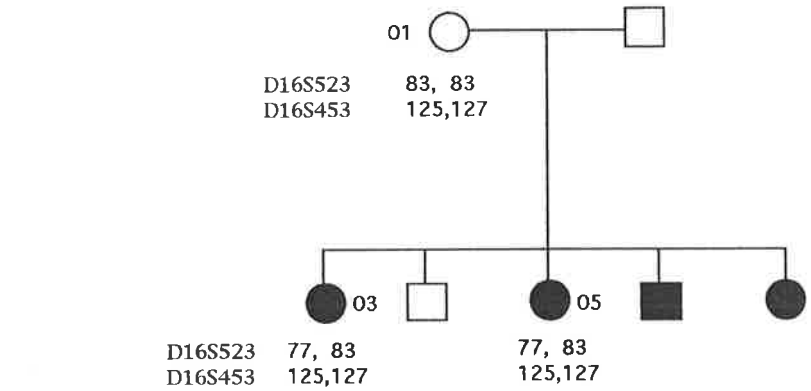

## Pedigree 34 (# 102)



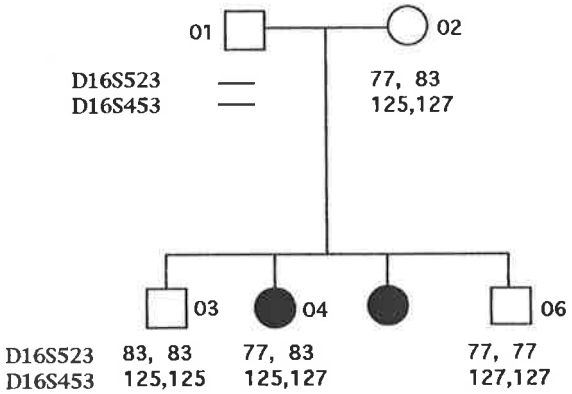
## Pedigree 35 (# 104)



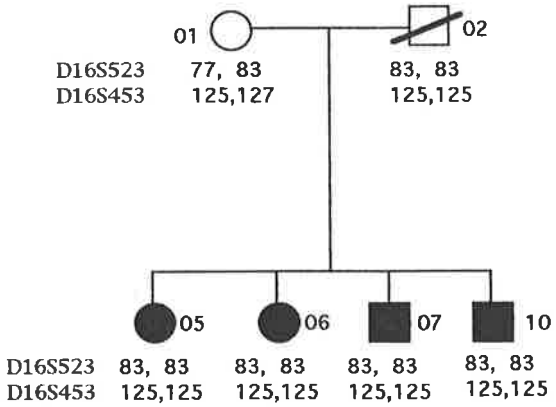

## Pedigree 36 (# 105)



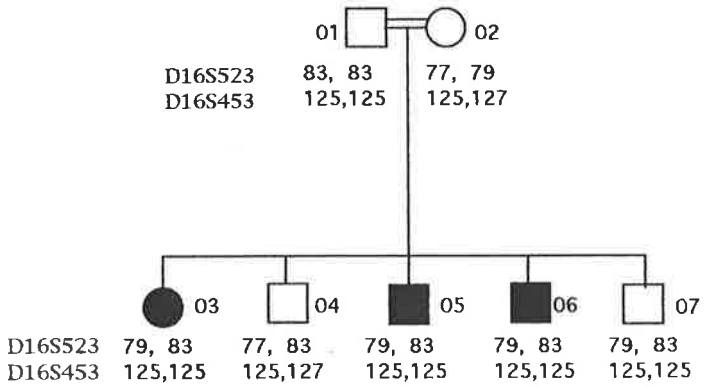
## Pedigree 37 (# 106)



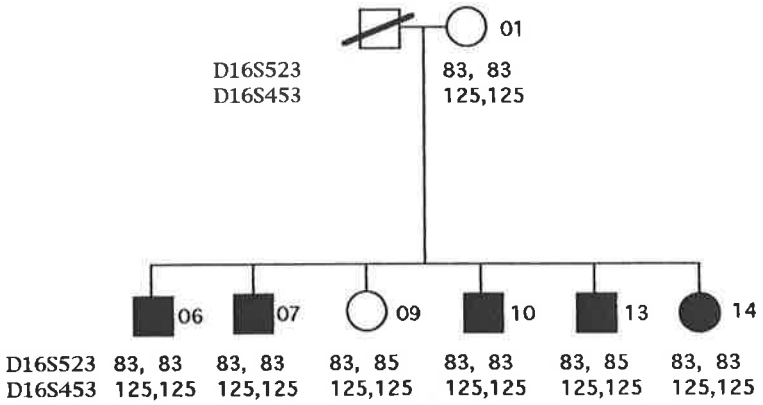

## Pedigree 38 (# 108)



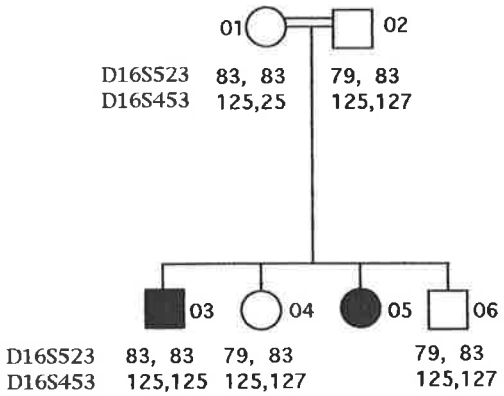
## Pedigree 39 (# 109)



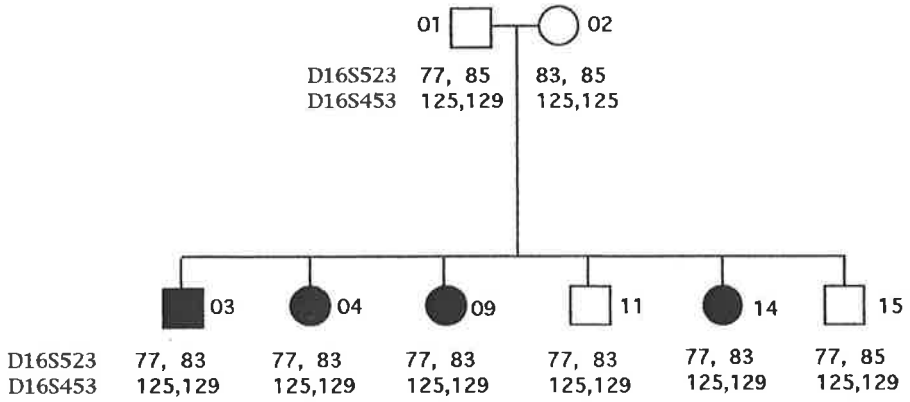
## Pedigree 40 (# 112)



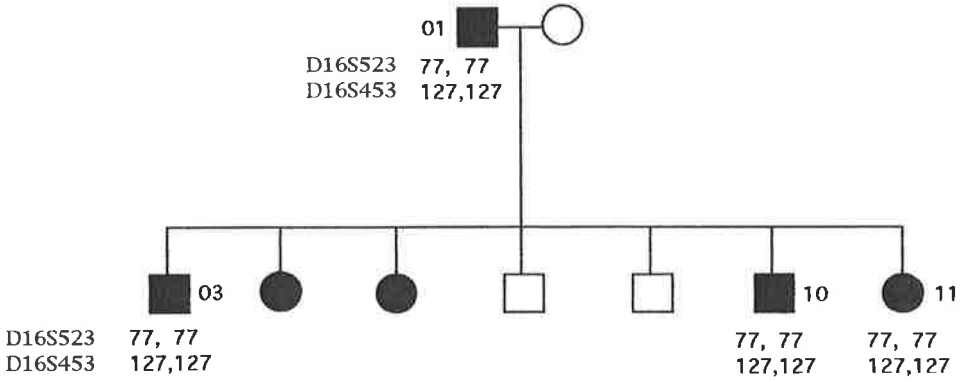

## Pedigree 41 (# 115)



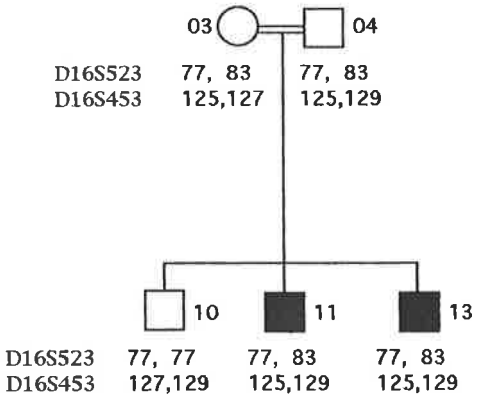
## Pedigree 42 (#117)



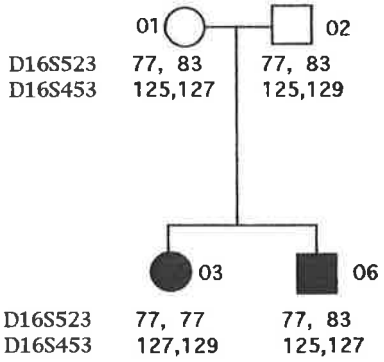

## Pedigree 43 (# 118)



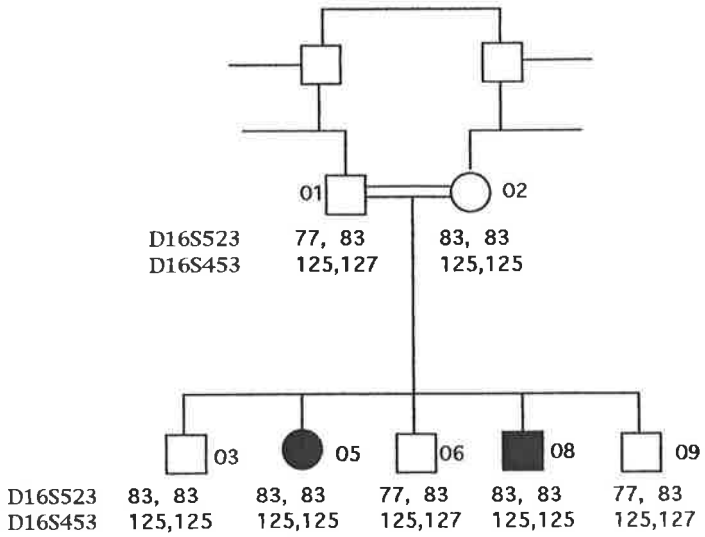
## Pedigree 44 (# 119)



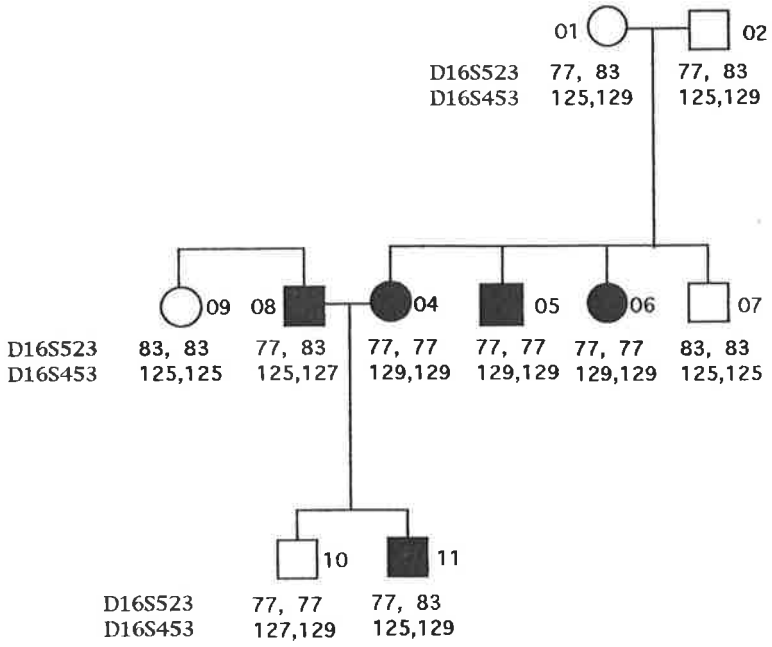

## Pedigree 45 (# 121)



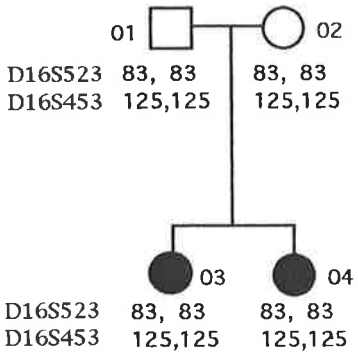
## Pedigree 46 (# 123)



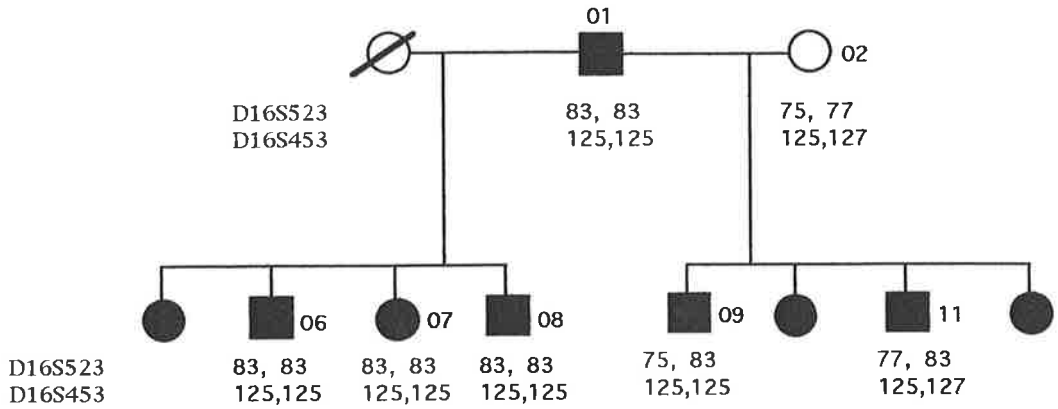

## Pedigree 47 (# 125)



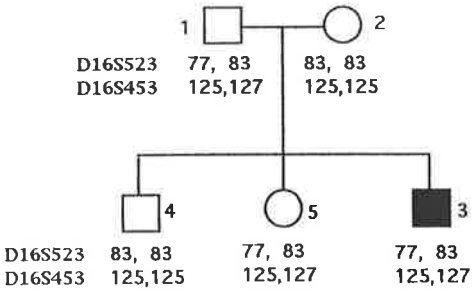
## Pedigree 48 (# 130)



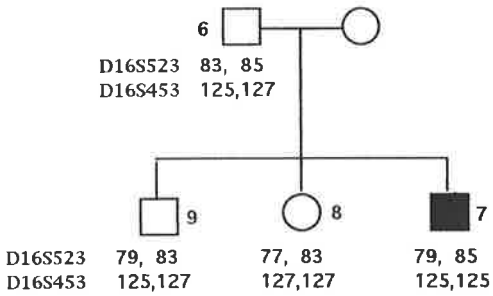

## Pedigree 49 (# 145)



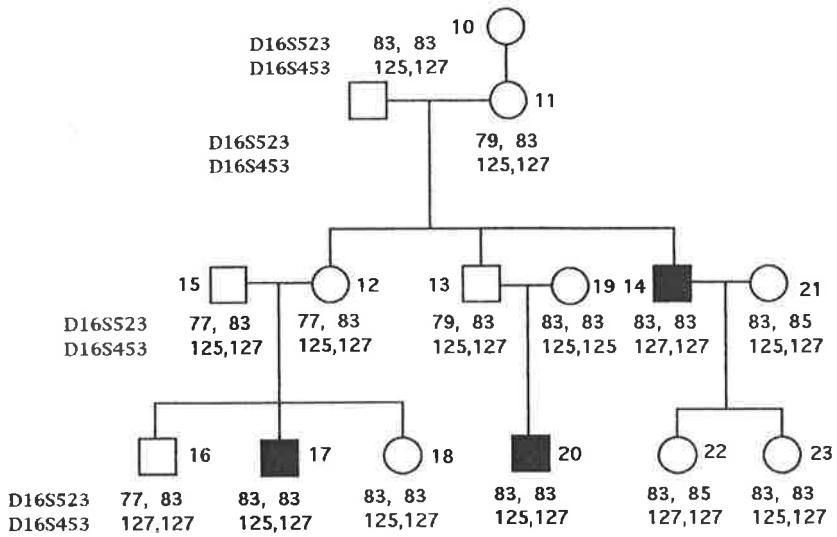
## Pedigree 50 (F1)



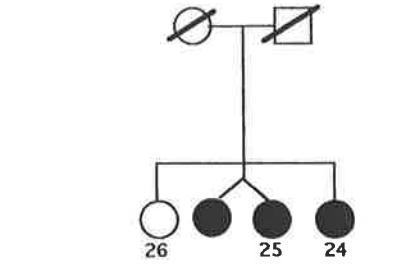

## Pedigree 51 (F2)



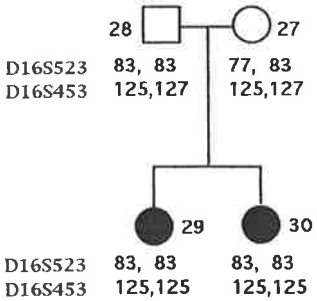

## Pedigree 52 (F3)



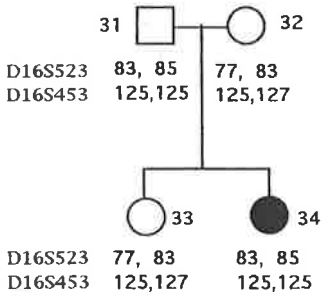
## Pedigree 53 (F4)



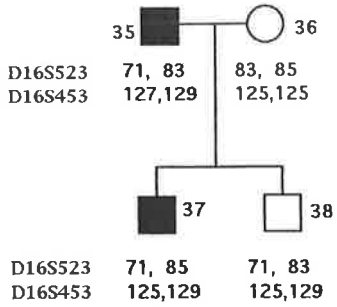

## Pedigree 54 (F5)



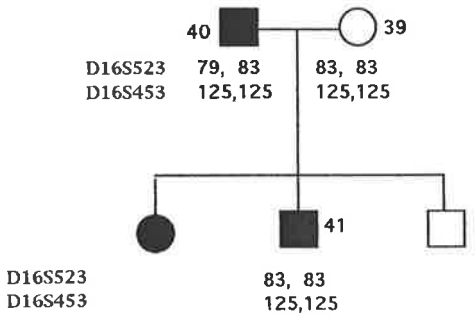

## Pedigree 55 (F6)



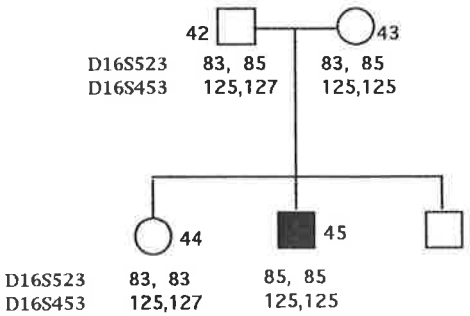
## Pedigree 56 (F7)



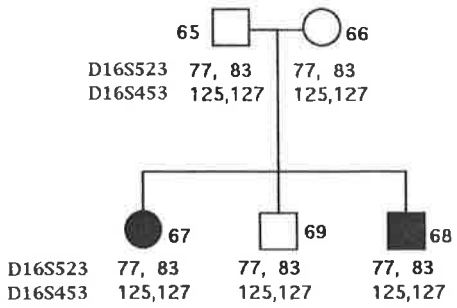

## Pedigree 57 (F9)



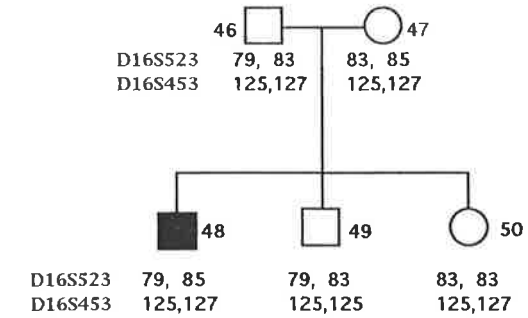

## Pedigree 58 (F10)



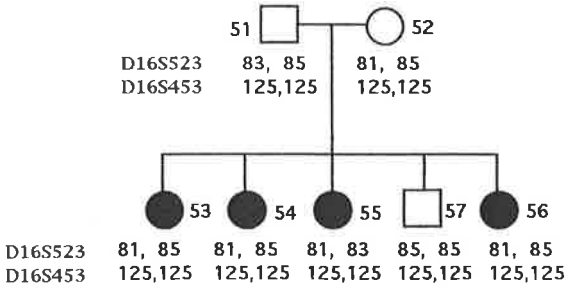

## Pedigree 59 (F11)



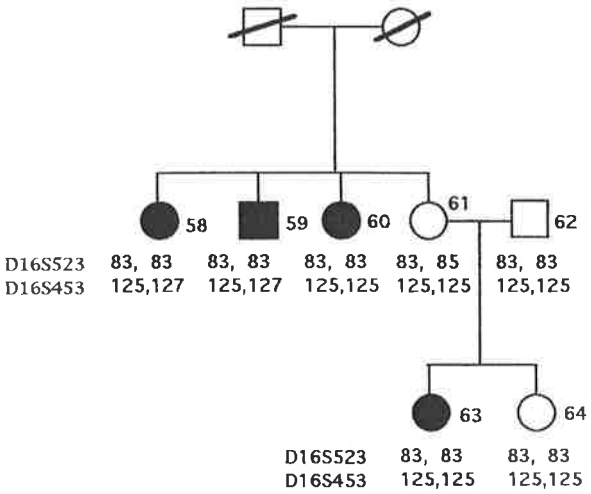
## Pedigree 60 (F12)




## Pedigree 61 (F13)




## Pedigree 62 (F14)



# APPENDIX III

## PUBLICATIONS

This appendix contains a list of papers that were published or are currently submitted for publication. In the publications which directly arose from this thesis, the roles of the candidate are specified; followed by copies of the papers and manuscripts.

1. Richards RI, Holman K, Shen Y, Kozman H, Harley H, Brook D and Shaw D (1991) Human glandular Kallikrein genes: genetic and physical mapping of the KLK1 locus using a highly polymorphic microsatellite PCR marker. *Genomics* 11:77-82

2. Richards RI, Shen Y, Holman K, Kozman H, Hyland VJ, Mulley JC and Sutherland GR (1991) Fragile X Syndrome: Diagnosis using highly polymorphic microsatellite markers. *Am J Hum Genet* 48:1051-1057.

3. Shen Y, Holman K, Thompson A, Kozman H, Callen DF, Sutherland GR and Richards RI (1991) Dinucleotide repeat polymorphism at the *D16S288* locus. *Nucl Acid Res* 19:5445

   The candidate isolated, physically mapped and characterized the (AC)n repeat marker *D16S288.*, and wrote the paper. (Chapters 2, 3 and 4)

4. Chen LZ, Shen Y, Holman K, Thompson A, Lane S, Richards RI, Sutherland GR and Callen DF (1991) An STS at the *D16S290* locus. *Nucl Acid Res* 19:5793.

5. Thompson AD, Shen Y, Holman K, Sutherland GR, Callen DF and Richards RI (1992) Isolation and characterisation of (AC)n microsatellite genetic markers from human chromosome 16. *Genomics* 13:402-408

   The candidate carried out some experiments on screening, sequencing and characterizing of (AC)n repeats. (Chapter 4)

6. Shen Y, Baker E, Callen DF, Sutherland GR, Willson TA, Rakar S and Gough NM(1992) Localization of the human GM-CSF receptor β chain gene (CSF2RB) to chromosome 22q12.2-22q13.1. *Cytogene Cell Genet* 61:175-177

7. Shen Y, Thompson AT, Holman K, Callen DF, Sutherland GR and Richards RI (1992) Four dinucleotide repeat polymorphisms on human chromosome 16 at *D16S289, D16S318, D16S319* and *D16S320. Hum Mol Genet* 1: 773

   The candidate isolated, physically mapped and characterized these four (AC)n repeat markers, and wrote the paper. (Chapters 2, 3 and 4)

8. Schuffenhauer S, Callen DF, Seidel H, Shen Y, Lederer G and Murken J (1992)*De novo* interstitial deletion 16 ( q12.1q13) of paternal origin in a 10 year old boy. *Clinic Genet* 42:246-250

9. Callen DF, Doggett NA, Stallings RL, Chen LZ, Whitmore SA, Lane SA, Nancarrow JK, Apostolou S, Thompson AD, Lapsys NM, Eyre HJ, Baker EG, Shen Y, Holman K, Phillips H, Richards RI and Sutherland GR (1992) High-resolution cytogenetic-based physical map of human chromosome 16. *Genomics* 13:1178-1185

   The candidate contributed 8 (AC)n repeat markers (*D16S531*, 16AC7.22, *D16S319, D16S288, D16S317, D16S320, D16S318* and *D16S289*) on the high-resolution cytogenetic-based physical map of human chromosome 16. (Chapter 4)

10. Callen DF, Thompson AD, Shen Y, Phillips HA, Richards RI, Mulley JC and Sutherland GR (1993) Incidence and origin of "Null" alleles in the $(AC)_n$ microsatellite markers from chromosome 16. *Am J Hum Genet* 52: 922-927

    The candidate contributed the genotype data of 6 (AC)n repeats (*D16S291, D16S94, D16S319, D16S301, D16S300* and *D16S289*). (Chapter 4)

11. Aksentijevich I, Pras E, Gruberg L, Shen Y, Holman K ,Helling S, Prosin L, Sutherland GR, Richards RI, Ramsurg M, Dean M, Pras M, Amos CI and

Kastner DL (1993) Refined mapping of the gene causing Familial Mediterranean Fever by linkage and homozygosity studies. *Am J Hum Genet* 53: 451-461

The candidate genotyped two (AC)n repeat markers (*D16S291* and *D16S94*)through 34 FMF families and contributed the genotype data (Chapter 6)

12. Aksentijevich I, Pras E, Gruberg L, Shen Y, Holman K ,Helling S, Prosin L, Sutherland GR, Richards RI, Dean M, Pras M and Kastner DL (1993) Familial Mediterranean Fever in Moroccan Jews: demonstration of a founder effect by extended haplotype analysis. *Am J Hum Genet* 53: 644-651

The candidate contributed the genotype data of (AC)n markers *D16S291* and *D16S94*. (Chapter 6)

13. Shen Y, Holman k, Doggett NA, Callen DF, Sutherland GR and Richards RI (1993a) Five dinucleotide repeat polymorphisms on human chromosome 16q24.2-q24.3.*Hum Mol Genet* 2: 1504

The candidate isolated, physically mapped and characterized these five (AC)n repeat markers (*D16S449, D16S393, D16S332, D16S392* and *D16S363*), and wrote the paper. (Chapters 2, 3 and 4)

14. Shen Y, Holman k, Doggett NA, Callen DF, Sutherland GR and Richards RI (1993b) Six dinucleotide repeat polymorphisms on human chromosome 16q12.1-q24.1. *Hum Mol Genet* 2: 1505

The candidate isolated, physically mapped and characterized these six (AC)n repeat markers (*D16S395, D16S450, D16S451, D16S389, D16S347*, and *D16S390*), and wrote the paper. (Chapters 2, 3 and 4)

15. Shen Y, Holman k, Doggett NA, Callen DF, Sutherland GR and Richards RI (1993c) Three dinucleotide repeat polymorphisms on human chromosome 16p13.11-p13.3. *Hum Mol Genet* 2: 1506

The candidate isolated, physically mapped and characterized these three (AC)n repeat markers (*D16S454, D16S453* and *D16S49*), and wrote the paper. (Chapters 2, 3 and 4).

16. Shen Y, Holman k, Doggett NA, Callen DF, Sutherland GR and Richards RI (1993d) Four dinucleotide repeat polymorphisms on human chromosome 16. *Hum Mol Genet* 2: 1745

    The candidate isolated, physically mapped and characterized these four (AC)n repeat markers (*D16S383, D16S452, D16S524* and *D16S523*), and wrote the paper. (Chapters 2, 3 and 4).

17. Callen DF, Eyre H, Lane S, Shen Y, Hansmann I, Spinner N, Zackai E, McDonald-McGinn, Schuffenhauer S, Wauters J, van Thienen M-N, van Roy B, Sutherland GR and Haan EA (1993) High resolution mapping of interstitial long arm deletions of chromosome 16: relationship to phenotype. *J Med Genet* 30: 828-832

18. Shen Y, Holman k, Doggett NA, Callen DF, Sutherland GR and Richards RI (1994a)Dinucleotide repeat polymorphisms at the *D16S525, D16S359, D16S531* and *D16S522* loci. *Hum Mol Genet* 3: 210

    The candidate isolated, physically mapped and characterized these four (AC)n repeat markers and wrote the paper. (Chapters 2, 3 and 4)

19. Shen Y, Kozman MH, Thompson A, Phillips HA, Holman K, Nancarrow J, Lane S, Chen L-Z, Apostolou S, Doggett N, Callen DF, Mulley JC, Sutherland GR and Richards RI (1994b) A PCR-based genetic linkage map of human chromosome 16. Submitted.

    The candidate documented the all data from isolation, characterization, physical and genetic mapping of 79 STR markers (27 (AC)n repeat markers isolated by the candidate, 20 (AC)n repeat markers isolated by other members of the candidate's laboratory and 32 STR markers isolated by other laboratories), and wrote the paper under Dr. R. I. Richards' supervision. (Chapters 2, 3 and 4)

GENOMICS 11, 77–82 (1991)

# Human Glandular Kallikrein Genes: Genetic and Physical Mapping of the *KLK1* Locus Using a Highly Polymorphic Microsatellite PCR Marker

ROBERT I. RICHARDS,* KATHERINE HOLMAN,* YANG SHEN,*,[1] HELEN KOZMAN,* HELEN HARLEY,† DAVID BROOK,† AND DUNCAN SHAW†

*Department of Cytogenetics and Molecular Genetics, Adelaide Children's Hospital, North Adelaide, Australia; and †Institute of Medical Genetics, University of Wales College of Medicine, Heath Park, Cardiff CF44XN, Wales, United Kingdom

Received November 15, 1990; revised May 6, 1991

**We describe a highly polymorphic microsatellite repeat sequence, *KLK1* AC, which is located 3′ to the human glandular kallikrein gene (*KLK1*) at 19q13.3–13.4. A multiplex PCR was developed to simultaneously genotype the *KLK1* AC repeat length polymorphism and a similar repeat at the adjacent *APOC2* locus at 19q13.2. Genotypes from these two loci in the 40 large kindred pedigrees from the Centre d'Etude du Polymorphisme Humain were used in conjunction with the background genetic map to establish a multipoint linkage map. The *KLK1* locus was also localized physically using somatic cell hybrid DNA templates for polymerase chain reaction analysis. Both genetic and physical mapping studies are consistent with the assignment cen–*APOC2*–*KLK1*–*D19522*–qter. The linkage map places *KLK1* approximately 10 cM distal to *APOC2*. These markers therefore flank the myotonic dystrophy gene and may be useful for diagnosis.** 

## INTRODUCTION

Genetic linkage mapping has, until recently, relied in large part upon restriction fragment length polymorphisms (RFLPs) to provide markers for pedigree analysis. These markers are not, in general, highly informative and are laborious to type. The advent of polymerase chain reaction (PCR) technology has allowed the analysis of length polymorphisms present in microsatellite sequences of the type $(AC \cdot GT)_n$ (Weber and May, 1989; Litt and Luty, 1989). There are approximately 50,000 such sequences in the human genome which appear to be randomly dispersed. The majority of these sequences exhibit length variation leading to heterozygosity values significantly higher than is common for RFLPs (Weber and May, 1989). There are two approaches to utilizing these sequences to establish genetic linkage. One is to isolate AC repeat sequences at random throughout the genome or for a specific chromosome (Luty *et al.*, 1990). A sufficient number of these markers randomly distributed will eventually allow linkage to be established by segregation in large affected pedigrees. The second approach is to target particular candidate genes for a disorder and characterize AC repeat sequences in close physical proximity for use as highly informative genetic markers. In this study we have characterized an AC repeat sequence in the vicinity of the human renal kallikrein (*KLK1*) gene (Sutherland *et al.*, 1988; Evans *et al.*, 1988), a candidate for genetic predisposition to hypertension. Oligodeoxyribonucleotide primers were designed from unique sequences flanking the microsatellite and used in PCR to type length variation at this locus. Reaction conditions were established for the simultaneous typing of the *KLK1* and *APOC2* AC repeats, allowing rapid analysis of these markers through 40 pedigrees from the Centre d'Etude du Polymorphisme Humain (CEPH). These data were then used to construct a multipoint linkage map of *KLK1* AC in relation to other genetic markers at 19q13.

## MATERIALS AND METHODS

Synthetic poly(AC · GT) was purchased from Pharmacia P.L. and radioactively labeled in a multiprime reaction (Amersham) using [$\alpha$-$^{32}$P]dCTP (3000 Ci/mmol, Amersham). This probe was then used to locate $(AC \cdot GT)_n$ sequences in the vicinity of *KLK1* on a Southern blot of various restriction enzyme digests of the plasmid pHRK8.8 (which contains the intact

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under Accession No. M65145.

[1] On leave from the International Peace Maternity and Child Health Hospital of China Welfare Institute, Shanghai, People's Republic of China.



0888-7543/91 $3.00

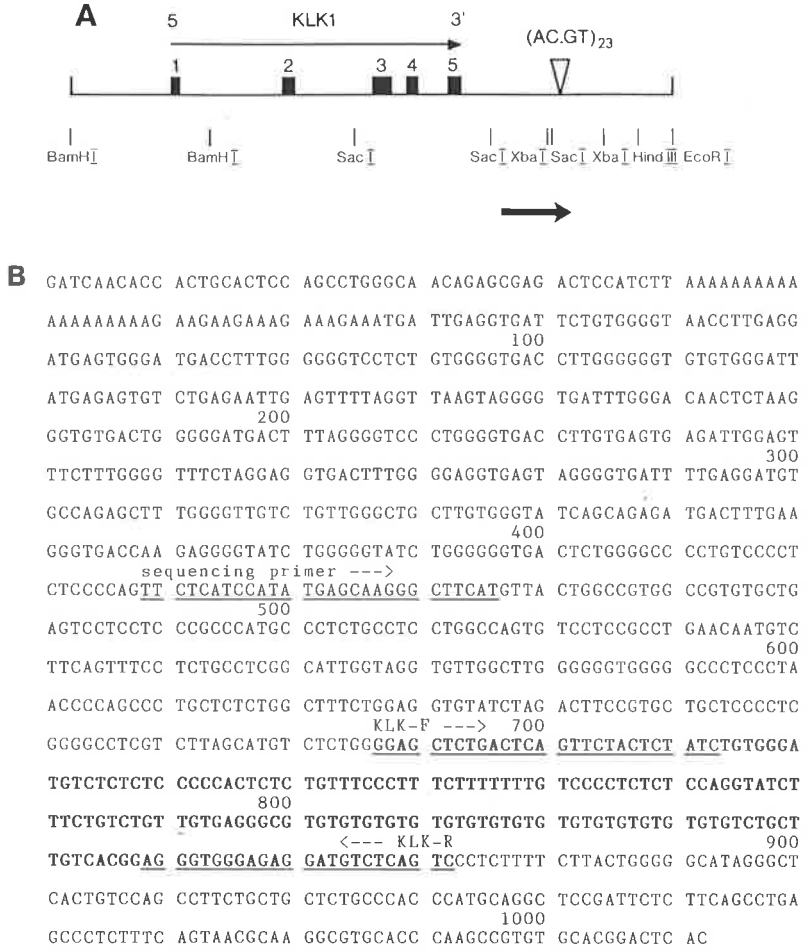



**FIG. 1.** Location and sequence of the *KLK1* AC repeat. (**A**) Physical map of pHRK8.8 indicating the exons and direction of transcription (thin arrow) of the *KLK1* gene and the location of the AC repeat. Orientation of the sequenced region (thick arrow) was determined by the position of *Xba*I (696) and *Sac*I (748) restriction endonuclease recognition sequences. (**B**) Sequence of the *Sau*3AI restriction fragment containing the *KLK1* AC repeat sequence. Also shown are the sequences used for a specific sequencing primer to extend the sequence to the AC repeat region and the sequences of the two primers used in PCR to type the AC repeat length polymorphism.

*KLK1* gene on 8.8 kb of genomic DNA) and to screen bacteriophage M13 subclones of a *Sau*3A digest of pHRK8.8. Positive clones were subjected to dideoxyribonucleotide sequence analysis (Sanger *et al.*, 1977) using initially a universal sequencing primer and subsequently a specific sequencing primer to determine $(AC)_n$ and flanking sequences. The location of the $(AC)_n$) microsatellite repeat and the deduced nucleotide sequence from the *Sau*3A subclone are illustrated in Fig. 1.

Oligodeoxyribonucleotide primers were designed from the deduced sequence and synthesized on an Applied Biosystems PCR-mate Synthesizer. These primers were located approximately 190 bases apart to allow simultaneous genotyping of the *APOC2* AC repeat that is also located at 19q13. The length of the previously characterized *APOC2* AC repeat PCR products range from 129 to 165 bases (Weber and May, 1989).

The *KLK1* AC and *APOC2* AC repeats were typed in a 10-$\mu$l polymerase chain reaction according to Kogan *et al.* (1987) and including 1 $\mu$Ci of [$\alpha$-$^{32}$P]dCTP. Reaction products were subjected to electrophoresis on 6% sequencing gels and autoradiography for 2 days using an intensifying screen.

Somatic cell hybrids used for physical localization included G35F3B (G35), which contains 19pter→q13.2, and GM89A99c7B (GM89), which contains 19q13.2→qter. G35 and GM89 were generated from a translocation of t(X;19) that breaks in the *CKMM* locus on 19q13.2 and therefore contain an entire human chromosome 19 between them. GM89 is a mouse cell hybrid. The third hybrid 20XP-3542-1-4 (20XP) contains a fragment of human chromosome

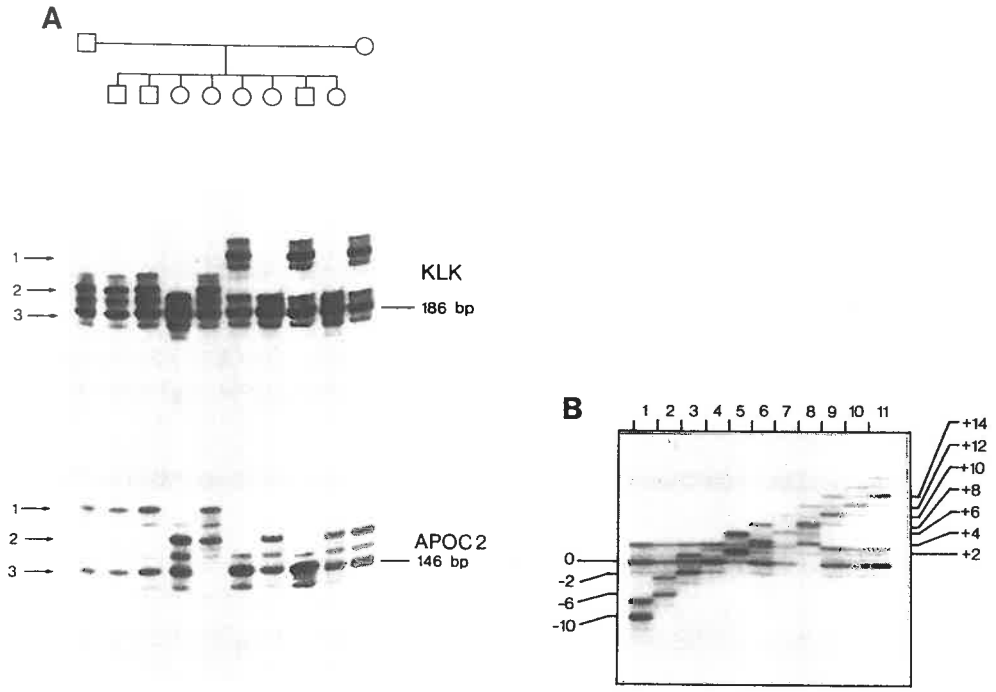



**FIG. 2.** Simultaneous typing of *KLK1* and *APOC2* AC repeats. **(A)** Autoradiograph of multiplex PCR of *KLK1* and *APOC2* AC repeat markers after electrophoresis on a 6% polyacrylamide gel. Members of CEPH family 35 are shown in the pedigree above the corresponding lanes of the gel. Both parents are heterozygous for both AC markers. *1, 2,* and *3* refer to the major products seen for each of the three alleles observed on each locus. The length of the most common allele (*3* for each locus) is given in base pairs. **(B)** Range of alleles at *KLK1* AC. Autoradiograph of *KLK1*AC PCRs after electrophoresis on a 6% polyacrylamide gel. Chromosomal DNA from individuals exhibiting multiple alleles of the AC repeat were genotyped. Assignments were as follows lane **1,** 0, −10; lane **2,** 0, −6; lane **3,** 0, −2; lane **4,** 0, 0; lane **5,** +2, +2; lane **6,** 0, +4; lane **7,** 0, +6; lane **8,** +4, +8; lane **9,** 0, +10; lane **10,** 0, +12; lane **11,** 0, +14.

19 that includes the *BCL3, APOC2, CKMM,* and *ERCC1* genes but not *D19S22* (Harley *et al.,* unpublished results) and is a CHO hybrid that was obtained from Dr. M. Siciliano (University of Texas).

## RESULTS

AC repeat sequences were identified at the *KLK1* locus by hybridization to restriction digests of cloned chromosomal DNA from the region (Evans *et al.,* 1988; Digby *et al.,* 1989) with radioactivity labeled poly(AC · GT). One of these sequences (pHRK8.8) contained the glandular kallikrein gene and the AC repeat containing sequence mapped approximately 1 kb from the 3′ end of the *KLK1* gene. The AC repeat and its flanking sequences were determined by subcloning a *Sau*3AI digest of pHRK8.8 into M13mp18. Positive clones were all found to have the insert in the same orientation with the repeat sequence greater than 600 bp from the sequence priming site. To reach the repeat sequence a specific sequencing primer was designed and synthesized corresponding to residues 488 to 516 (Fig. 1). Sequencing from this priming site enabled the repeat sequence and its flanking sequences to be determined. Forward and reverse oligodeoxyribonucleotide primers suitable for PCR determination of length polymorphism of the AC repeat were designed, optimizing for GC composition and

**TABLE 1**

***KLK1* Allele Frequency (124 Unrelated Individuals)**

| Allele | Frequency |
|---|---|
| −10 | Not observed |
| −8 | Not observed |
| −6 | 0.004 |
| −4 | 0.004 |
| −2 | Not observed |
| 0 | 0.540 |
| +2 | 0.089 |
| +4 | 0.202 |
| +6 | 0.016 |
| +8 | 0.024 |
| +10 | 0.036 |
| +12 | 0.044 |
| +14 | 0.036 |
| +16 | 0.004 |

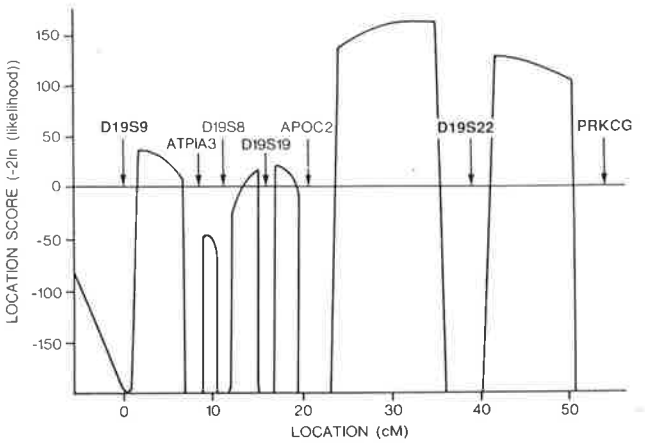



FIG. 3. Computed location of the *KLK1* locus. Plot of the probability of the location of the *KLK1* locus in relation to other genetic markers at 19q13.

product length so as to allow for simultaneous genotyping of the *KLK1* AC and adjacent *APOC2* AC repeats (Fig. 2). The PCR conditions of Kogan *et al.* (1987) were utilized as these have been successfully applied to multiplex reactions (Chamberlain *et al.*, 1988). As with other AC repeat genotypes, multiple bands were observed for each allele (Fig. 2). One of these additional bands is most likely the GT-containing strand (with reduced autoradiographic signal due to lower incorporation of [$\alpha$-$^{32}$P]dCTP than its AC containing counterpart). The origin of additional faint bands at each allele is not known although they are a common feature of these repeats (Weber and May, 1989; Litt and Luty, 1989).

The simultaneous typing of the *APOC2* and *KLK1* AC repeats allowed rapid analysis of both loci through the CEPH pedigrees. An example of this analysis (CEPH family 35) is shown in Fig. 2 as is a range of the alleles exhibited by this locus. The allele frequencies for unrelated individuals in the CEPH pedigrees are shown in Table 1. Certain rare alleles, including some of those illustrated in Fig. 2, have only been observed in further population studies (Richards and Soubrier, unpublished). The high heterozygosity of these AC markers results in 36 of the 40 CEPH families being informative for *KLK1* AC and only 2 of these were uninformative for *APOC2* AC. In addition the multiple alleles at each locus mean that the information content is frequently retained when both parents are heterozygous. All of these factors combine to give extraordinarily high probabilities for the resultant linkage map, for example, the odds against inversion of the *APOC2* and *KLK1* loci are $6 \times 10^{35}$ to 1.

The multipoint linkage map of *KLK1* AC in relation to other genetic markers at 19q13 was performed using the LINKAGE (version 4.9) computer program (Lathrop *et al.*, 1984). A plot of the location of *KLK1* AC on the background genetic map is shown in Fig. 3, as are the order of loci and the odds against inversion of adjacent loci in Table 2.

Physical mapping of the *KLK1* locus to human chromosome 19 somatic cell hybrids was also undertaken using the *KLK1* AC PCR marker with the *APOC2* AC PCR marker as an internal control. The cell lines GM89 and G35 are reciprocal translocations of t(X;19) with a breakpoint at the *CKMM* locus (Brook *et al.*, 1984, 1991), while 20XP has a breakpoint distal to *ERCC1* but does not contain D19S22 (Stallings *et al.*, 1988, and unpublished results). Products from the *KLK1* locus were present in the hybrid GM89 and absent in the hybrids G35 and 20XP, consistent with the genetic localization (Fig. 4).

## DISCUSSION

The majority of RFLPs have relatively low information contents, providing only low probabilities that a pedigree will be simultaneously informative at two loci. Length polymorphisms observed for sequences of the type $(AC \cdot GT)_n$, where $n$ is $\geq 15$, have polymorphism information content values usually in the range 0.5 to 0.8, making them very attractive for use as genetic markers (Weber and May, 1989). In addition their frequency in the human genome (~50,000 copies) supports their general application as genetic markers, providing that they are indeed randomly distributed.

The glandular (renal) kallikrein gene is one of a number of candidates as sites for mutation leading to genetic predisposition to hypertension (Berry *et al.*, 1989). Its gene product is an enzyme responsible for the generation of kinins which are potent vasodilators. In an effort to identify the gene(s) responsible for the genetic component of hypertension we have begun to characterize highly polymorphic markers in the vicinity of such candidate genes. These markers,

**TABLE 2**

**Genetic Linkage Map of *KLK1* AC**

Order of loci

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| $\theta_m$ | 0.07 | 0.023 | 0.036 | 0.034 | 0.097 | 0.052 | 0.102 |
| pIJ2— | ATP1A3— | D19S8— | D19S19— | AP0C2— | KLK1— | D19S22— | PKCG |
| $\theta_f$ | 0.082 | 0.027 | 0.043 | 0.04 | 0.112 | 0.061 | 0.118 |
| $Z$ | 8.7 | 22.7 | 4.25 | 9.62 | 28.76 | 11.7 | 7.3 |

Sex difference 1.18

Inversion of adjacent loci

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | $1 \times 10^6$ | | 290 | | $6 \times 10^{35}$ | | $1 \times 10^8$ |
| pIJ2— | ATP1A3— | D19S8— | D19S19— | AP0C2— | KLK1— | D19S22— | PKCG |
| | | 650 | | 306 | | $1 \times 10^5$ | |

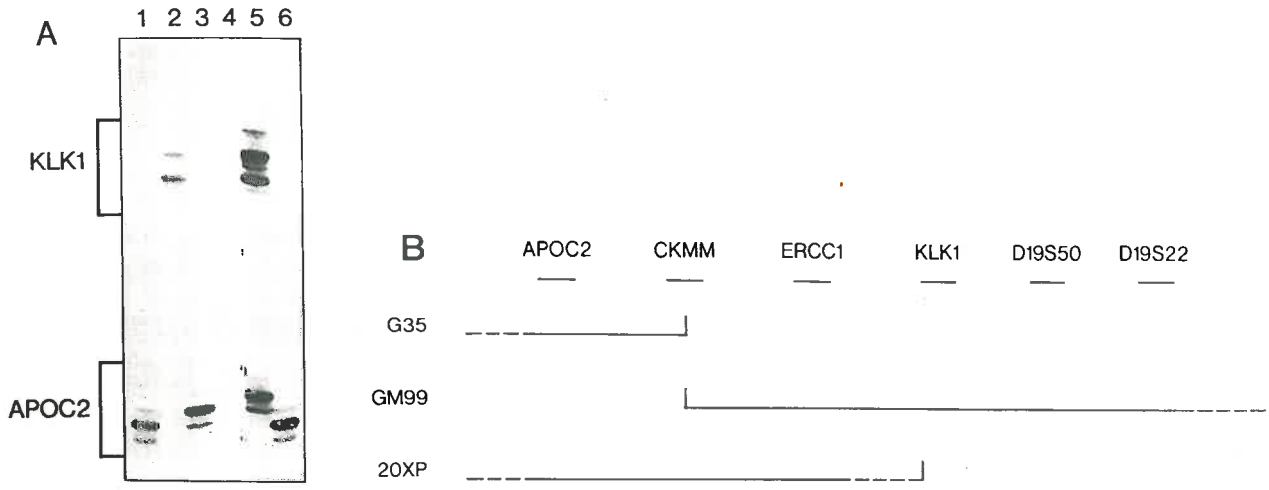



FIG. 4. Physical localization of the *KLK1* locus. (A) PCR products from the *KLK1* and *APOC2* AC repeats using somatic cell hybrid DNAs as templates. Lanes 1 and 6, G35; lane 2, GM89; lane 3, 20XP; lane 4, mouse; lane 5, human. (B) Portions of human chromosome 19 contained in somatic cell hybrids used to physically locate *KLK1* AC sequences. The location of other 19q13 loci in relation to the chromosome 19 breakpoints is also shown.

being more informative than RFLPs, will increase the probability of observing linkage between the disorder and the genetic marker in an affected pedigree. Such pedigrees have now been described by two laboratories (Berry *et al.*, 1989; Cambien *et al.*, 1988) and their analysis is in progress.

The *APOC2* AC repeat has previously been described (Taylor *et al.*, 1989) as a diagnostic marker suitable for linkage analysis of myotonic dystrophy (*DM*). The *KLK1* and *APOC2* AC repeats flank the *DM* locus and therefore the simultaneous typing of these repeats ought to provide a rapid and accurate means of following the mutant *DM* allele through affected families. Such a clinical application will be described elsewhere.

Physical mapping of sequences to the long arm of human chromosome 19 has been mainly by *in situ* hybridization. In some cases somatic cell hybrids with breakpoints in this region have been used in conjunction with *in situ* hybridization, allowing the order of some loci to be determined. This region of the human genome is not frequently involved in translocations and so the number of breakpoints that divide up the region is relatively small (Schonk *et al.*, 1990). Since the gene for myotonic dystrophy is located at 19q13.2–q13.3 there has been intense interest in both physical and genetic maps of the region and a large number of genes and anonymous DNA fragments have been mapped. Where restriction fragment length polymorphisms have been identified using these sequences as hybridization probes then their genetic location and hence physical order have been established. (Le Beau *et al.*, 1989; Ropers and Pericak-Vance, 1990).

The band at human chromosome 19q13 contains nine loci that map to a region of synteny on mouse chromosome 7. Analysis of such syntenic regions has been suggested as a means to identify specific disease genes and in this particular case the gene responsible for myotonic dystrophy. For such an approach to be successful it is essential that the order of the genes in the respective syntenic regions be maintained. The human genetic mapping of the *KLK1* AC repeat places this locus between *ERCC* and *PRKCG*. Genetic mapping of the mouse kallikrein gene family, which includes *Ngfg*, places the mouse locus outside of, and distal to, this interval in the vicinity of the *Lhb* (Saunders and Seldin, 1990). The human *LHB* locus also maps outside and distal to the *ERCC–PRKCG* interval and therefore while the clustering of analogous genes to syntenic regions is frequently observed conservation of the order of these genes may well be lost due to inversion events.

The general application of microsatellite markers shows a great deal of promise, providing that they are randomly distributed, to enable the establishment of a 1-cM map of the entire human genome. While analysis of affected pedigrees for polymorphisms associated with candidate disease genes may only succeed in eliminating these loci as sites for genetic predisposition to the disorder, the ultimate "saturation" of the human genetic map will succeed in establishing the site(s) of such lesions should a true heritable basis for the disorder exist.

## ACKNOWLEDGMENTS

Robert I. Richards thanks Shelley Richards for support and encouragement during the course of this work. Yang Shen was the

recipient of a World Health Organization Fellowship. We are grateful to Drs. Grant Sutherland and John Mulley for constructive advice on the project and the manuscript and thank Dr. Michael Siciliano (Department of Genetics, M. D. Anderson Hospital and Tumor Institute, University of Texas System Cancer Centre, Houston TX 77030) for the 20XP hybrid cell line and the Adelaide Children's Hospital Research Fund for financial support.

## REFERENCES

1. Berry, T. D., Hasstedt, S. J., Hunt, S. C., Wu, L. L., Smith, J. B., Ash, K. O., Kuida, H., and Williams, R. R. (1989). A gene for high urinary kallikrein may protect against hypertension in Utah kindreds. *Hypertension* **13:** 3–8.
2. Brook, D., Shaw, D. J., Meredith, L., Bruns, G. A. P., and Harper, P. S. (1984). Localization of genetic markers and orientation of the linkage group on chromosome 19. *Hum. Genet.* **68:** 282–285.
3. Brook, J. D., Knight, S., Roberts, S. H., Harley, H. G., Walsh, K. V., Rundle, S. A., Freyne, K., Koch, M. C., Epstein, N. D., Wieringa, B., Schonk, D., Smeets, H., Haddingham, K., Siciliano, M. J., Palmer, D. K., Miles, J. S., Wolf, C. R., Fonatsch, C., and Shaw, D. J. (1991). The physical map of chromosome arm 19q: Some new assignments, confirmations and reassignments. *Hum. Genet.* **87:** 65–72.
4. Cambien, F., Alhenc-Gelas, F., Herbeth, B., Andre, J. L., Rakotovao, R., Gonzales, M. F., Allegrini, J., and Bloch, C. (1988). Familial resemblance of plasma angiotensin-converting enzyme level: The Nancy study. *Am. J. Hum. Genet.* **43:** 774–780.
5. Chamberlain, J. S., Gibbs, R. A., Ranier, J. E., Nguyen, P. N., and Caskey, C. T. (1988). Deletion screening of the Duchenne muscular dystrophy locus via multiplex DNA amplification. *Nucleic Acids Res.* **17:** 11,141–11,156.
6. Digby, M., Zhang, X-Y., and Richards, R. I. (1989). Human Prostate Specific Antigen (PSA) gene: Structure and linkage to the kallikrein-like gene, hGK-1. *Nucleic Acids Res.* **17,** 2137.
7. Evans, B. A., Zhang, X. Y., Close, J. A., Tregear, G. W., Kitamura, N., Nakanishi, S., Callen, D. F., Baker, E., Hyland, V. J., Sutherland, G. R., and Richards, R. I. (1988). Structure and chromosomal localization of the human renal kallikrein gene. *Biochemistry* **27:** 3124–3219.
8. Kogan, S. C., Doherty, M., and Gitschier, J. (1987). An improved method for prenatal diagnosis of genetic diseases by analysis of amplified DNA sequences. *N. Engl. J. Med.* **317:** 985–990.
9. Lathrop, G. M., Lalouel, J-M., Julier, C., and Ott, J. (1984). Strategies for multi-locus linkage analysis in humans. *Proc. Natl. Acad. Sci. USA* **81:** 3443–3446.
10. Le Beau, M. M., Ryan, D., Jr., and Pericak-Vance, M. A. (1989). Report of the committee on the genetic constitution of chromosomes 18 and 19. *Cytogenet. Cell Genet.* **51:** 338–357.
11. Litt, M., and Luty, J. A. (1989). A hypervariable microsatellite revealed by *in vitro* amplification of a dinucleotide repeat within the cardiac muscle actin gene. *Am. J. Hum. Genet.* **44:** 397–401.
12. Luty, J. A., Gupo, Z., Willard, H. F., Ledbetter, D. H., Ledbetter, S., and Litt, M. (1990). Five polymorphic microsatellite VNTRs on the human X chromosome. *Am. J. Hum. Genet.* **46:** 776–783.
13. Ropers, H. H., and Pericak-Vance, M. A. (1990). Report on the committee on the genetic constitution of chromosome 19. *Cytogenet. Cell Genet.* **55:** 218–228.
14. Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74:** 5463–5467.
15. Saunders, A. M., and Seldin, M. F. (1990). The Syntenic Relationship of Proximal Mouse Chromosome 7 and the Myotonic Dystrophy Gene Region on Human Chromosome 19q. *Genomics* **6:** 324–332.
16. Schonk, D., van Dijk, P., Riegmann, P., Trapman, J., Holm, C., Willcocks, T. C., Sillekens, P., van Verrooij, W., Wimmer, E., Guerts van Kessel, A., Ropers, H-H., and Wieringa, B. (1990). Assignment of seven genes to distinct intervals on the midportion of human chromosome 19q surrounding the myotonic dystrophy region. *Cytogenet. Cell Genet.* **54:** 15–19.
17. Stallings, R. L., Olson, E., Strauss, A. W., Thompson, L. H., Bachinski, L. L., and Siciliano, M. J. (1988). Human creatine kinase genes on chromosomes 15 and 19, and proximity of the gene for the muscle form to the genes for apolipoprotein C2 and excision repair. *Am. J. Hum. Genet.* **43:** 144–151.
18. Sutherland, G. R., Baker, E., Hyland, V. J., Callen, D. F., Close, J. A., Tregear, G. W., Evans, B. A., and Richards, R. I. (1988). Human prostate-specific antigen (APS) is a member of the glandular kallikrein gene family at 19q13. *Cytogenet. Cell Genet.* **48:** 205–207.
19. Taylor, G., Noble, J. S., Hall, J. L., Stewart, A. D., and Mueller, R. F. (1989). Hypervariable microsatellite for genetic diagnosis. *Lancet,* 454.
20. Weber, J. L., and May, P. E. (1989). Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am. J. Hum. Genet.* **44:** 388–396.

*Am. J. Hum. Genet. 48:1051–1057, 1991*

# Fragile X Syndrome: Diagnosis Using Highly Polymorphic Microsatellite Markers

Robert I. Richards, Yang Shen,[1] Katherine Holman, Helen Kozman, Valentine J. Hyland,[2] John C. Mulley, and Grant R. Sutherland

Department of Cytogenetics and Molecular Genetics, Adelaide Children's Hospital, North Adelaide

## Summary

**We describe two highly polymorphic microsatellite AC repeat sequences, VK23AC and VK14AC, which are closely linked to the fragile X at Xq27.3. Both VK23AC (*DXS297*) and VK14AC (*DXS292*) are proximal to the fragile site. Two-point linkage analysis in 31 fragile X families gave (*a*) a recombination frequency of 1% (range 0.00%–4%) with a maximum lod score of 32.04 for *DXS297* and (*b*) a recombination frequency of 7% (range of 3%–15%) with a maximum lod score of 12.87 for *DXS292*. Both of these polymorphisms are applicable to diagnosis by linkage in families with fragile X syndrome. A multipoint linkage map of genetic markers at Xq27.3 was constructed from genotyping these polymorphisms in the CEPH pedigrees. The *DXS292* marker is in the *DXS98-DXS297* interval and is 3 cM proximal to *DXS297*.**

## Introduction

The rare fragile site at Xq27.3 (*FRAXA*) is associated with the most common familial form of mental retardation (Sutherland and Hecht 1985). Prenatal diagnosis and carrier detection can be performed cytogenetically; however, incomplete penetrance of the fragile site renders this technology, in isolation, inaccurate. Diagnosis for individuals who do not express the fragile site therefore relies on polymorphic DNA markers closely linked to it (Sutherland and Mulley 1990).

The majority of DNA polymorphisms currently used for risk estimation are detected by Southern blot analysis of RFLPs (Suthers et al. 1991*a*). These markers are less common on the X than on the autosomes (Hofker et al. 1986), and laboratory analysis is labor intensive. In addition, the majority have only two alleles and therefore a maximum heterozygosity—and, in the case of X-linked markers, PIC—of 50%.

Recently, two laboratories have described polymorphisms associated with length variation in dinucleotide microsatellite repeats (Litt and Luty 1989; Weber and May 1989). These polymorphisms can be rapidly typed and are usually highly informative (compared with RFLPs). They are quite common; there are about 50,000 copies of the (AC)n repeat in the mammalian genome—i.e., approximately one every 50 kbp, if one assumes uniform distribution.

We have screened cloned DNA from the vicinity of Xq27.3 for dinucleotide microsatellite sequences to identify polymorphisms useful for linkage analysis near *FRAXA*. The established physical order for markers proximal to *FRAXA* is *DXS98* (4D-8)–*DXS292* (VK14)–*DXS369*(RN1)–*DXS297*(VK23)–*FRAXA* (Suthers et al. 1990). Two polymorphisms, VK23AC and VK14AC, are described, together with both their genetic mapping in the Centre d'Etude du Polymorphisme Humain (CEPH) pedigrees and the results of their application to 31 fragile X syndrome pedigrees.

## Material and Methods

Synthetic poly (AC.GT) (Pharmacia) was radioac-

Received November 29, 1990; revision received January 24, 1991.

Address for correspondence and reprints: Dr. R. I. Richards, Department of Cytogenetics and Molecular Genetics, Adelaide Children's Hospital, North Adelaide, South Australia 5006.

1. On leave from the International Peace Maternity and Child Health Hospital of China Welfare Institute, Shanghai.

2. Present address: Kolling Institute of Medical Research, Royal North Shore Hospital, Saint Leonards, New South Wales, Australia.


0002-9297/91/4806-0004$02.00

tively labeled with alpha-$^{32}$P-dCTP in a random-primed reaction (Multiprime; Amersham). AC repeat–containing DNA sequences were identified by hybridization to this probe in 0.5 M sodium phosphate pH 7.0, 7% SDS (without carrier DNA) at 65°C for 16 h and by washing at 65°C for 1 h in 2 × SSC. Clones tested constituted the VK series which had been mapped to Xq26-qter (Hyland et al. 1989).

DNA from positive lambda clones was digested with *Sau*3A and was subcloned into *Bam*HI-cut M13 mp18 for sequence analysis. Synthetic oligodeoxyribonucleotide primers suitable for PCR were designed from apparently unique sequences flanking the microsatellite AC repeats. Length polymorphism of the AC repeats was typed in a PCR using the reaction conditions of Kogan et al. (1987), except for the addition of the 1 μCi of alpha-$^{32}$P-dCTP to each reaction. These PCR conditions were used because they have been successfully applied to multiplex PCR (Chamberlain et al. 1988, Richards et al., submitted).

PCR incubations were performed in 10-μl volumes in a Perkin Elmer-Cetus thermal cycler for 10 cycles at 94°C for 60 s, at 60°C for 90 s, and at 72°C for 90 s, followed by 25 cycles at 94°C for 60 s, at 55°C for 90 s, and at 72°C for 90 s. The volume was adjusted to 40 μl with formamide loading buffer (95% formamide, 1 mM EDTA, 0.01% bromophenol blue, 0.01% xylene cyanol). After denaturation at 90°C for 3 min, 2.5-μl aliquots of each reaction mixture were subjected to electrophoresis in 6% polyacrylamide denaturing (7 M urea) gels. Genotypes were determined after autoradiography for 18–48 h.

Linkage analysis were based on 31 kindreds. Five small kindreds with isolated cases of fragile X were excluded from the analysis because of uncertainty about relative mutation rates in males and females. Parameters used were as given elsewhere (Suthers et al. 1991*a*), except for allele frequencies of the marker loci, VK23AC and VK14AC, frequencies which were determined from unrelated individuals in the informative fragile X families and are given in table 1.

Modification of the phenotypic coding of family members was made in those cases in which closely linked flanking markers clearly indicated carrier status different from that determined by the conventional definition of phenotype. In previous studies the phenotype has been defined only by cytogenetic expression of the fragile X and by mental retardation. Now that there are many markers closely linked on either side of *FRAXA* (Suthers et al. 1991*a*), the definition of phenotype for the purposes of the present study was extended to include the genotypes of closely linked surrounding markers. If a cytogenetically negative nonretarded individual has marker alleles both proximal and distal to *FRAXA*, clearly demonstrating that the individual carries the fragile X chromosome (if the possibility of double crossover is disregarded), then that individual will contribute more information to a linkage analysis if he or she is coded as a carrier than if he or she is coded as unaffected in a defined penetrance class. Conversely, if it could be demonstrated that such an individual does not carry the fragile X chromosome (if the possibility of double crossover is disregarded), then that individual also will contribute more information to a linkage analysis if he or she could be coded as a definite noncarrier. Determination of such carrier status by DNA markers does not rely on map distances estimated in the present study by using VK23AC and VK14AC. The genetic map distances on which carrier status is based were previously established in an independent analysis of linkage data (Suthers et al. 1991*a*). For the present study, first the unaffected individuals of either sex were coded as if they were affected, when closely linked informative markers flanking the fragile X demonstrated that they were carriers. Second, a penetrance class of 100% was assigned to unaffected individuals of either sex in whom closely linked informative markers flanking the fragile X demonstrated that they were not carriers. This both removed the option, available under incomplete pene-

**Table 1**

**Alleles, Allele Frequency, and Heterozygosity for VK23AC and VK14AC in Fragile X Pedigrees**

| Marker and Allele | No. of Chromosomes | Frequency | Heterozygosity |
|---|---|---|---|
| VK23AC: | 65 (60) | | .74 (.67) |
| +4 | | .00 (.02) | |
| +2 | | .08 (.03) | |
| 0 | | .34 (.53) | |
| −2 | | .32 (.13) | |
| −4 | | .11 (.17) | |
| −6 | | .15 (.12) | |
| −12 | | .00 (.02) | |
| VK14AC: | 44 (60) | | .53 (.58) |
| +6 | | .00 (.02) | |
| +4 | | .00 (.05) | |
| +2 | | .02 (.13) | |
| 0 | | .64 (.62) | |
| −2 | | .23 (.08) | |
| −4 | | .11 (.10) | |

NOTE.—Data in parentheses refer to CEPH pedigrees.

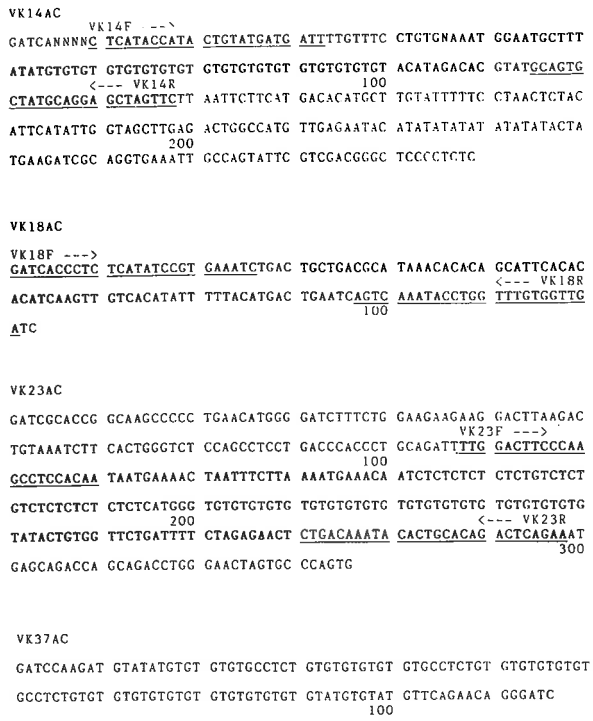



**Figure 1** AC/GT repeat sequences determined from VK14, VK18, VK23, and VK37. Nucleotide sequences containing AC.GT repeats from each of the markers near Xq27.3 are shown. Sequences used to design primers for genotyping polymorphism of repeat length are underlined. PCR products are shown in boldface.

trance, that they might carry the fragile X and allowed full utilization of data from these individuals.

Twenty-seven and 14 of the 31 fragile X families were at least partially informative for VK23AC and VK14AC, respectively. One additional family was informative for the VK23B *Hin*dIII RFLP but not for VK23AC. The analysis of fragile X families was carried out by using the LINKAGE (version 5.04) package of computer programs and incorporated the data from the *Hin*dIII RFLP of VK23B from the earlier study of Suthers et al. (1991*a*). Confidence intervals were obtained by the lod – 1 method (Conneally et al. 1985). Multipoint analysis was based on 40 normal families from CEPH (Dausset et al. 1990). Analyses of the CEPH families were carried out by using the LINKAGE (version 4.9) package for use with CEPH three-generation families.

## Results

### *Identification of (AC.GT) in Repeats at Xq27.3*

Nineteen human genomic DNA clones (lambda VK7, 9–11, 14, 16–18, 21, 23–25, 29, 34, 37, 40, 41, 44, and 47) which map to the interval Xq26-qter (Hyland et al. 1989) were screened for the presence of AC-repeat microsatellite sequences, and nine of these were positive. The four positive clones VK14, VK18, VK23, and VK37, which physically map closest to *FRAXA* (Suthers et al. 1990) were sequenced to determine both the length of AC repeats and the composition of unique flanking sequences. The relevant sequences from each of these regions are shown in figure 1, as are the location and sequences of PCR primers designed from them to type the AC-repeat-length polymorphism. No primers could be designed from the VK37 sequence, because of the close proximity between the *Sau*3AI cloning sites and the AC repeat.

### *Characterization of AC-Repeat Microsatellite Polymorphism*

Primers were specifically designed at distances sufficient to allow simultaneous typing of both VK23AC and VK14AC. Twenty unrelated individuals were typed for length variation of the AC repeats at the VK14, VK18, and VK23 loci. The VK18 repeat showed no polymorphism and was therefore not analyzed further. Both the VK14 and VK23 AC repeats revealed length variation inherited in a Mendelian fashion (fig. 2) in fragile X families and had heterozygosities greater than 50% (table 1). Key carriers from all available fragile X–linked mental retardation pedigrees were genotyped, and informative families were completed. Heterozygosities for unrelated individuals in the fragile X pedigrees were 53% VK14, and 74% VK23, compared with the 58% VK14 and 67% VK23 for unrelated individuals from unaffected (CEPH) pedigrees.

### *Two-Point Linkage Analysis*

Genotypes of AC-repeat length at the *DXS297* and *DXS292* loci were used to calculate genetic distance from *FRAXA* (table 2). *DXS292* was found to have a peak lod score of 12.87 at a recombination fraction of .07 (confidence interval .03–.15), and *DXS297* had a peak lod score of 32.04 at a recombination fraction of .01 (confidence interval .00–.04).

The lod score for the *FRAXA:DXS297* comparison remains positive at a recombination fraction of zero, despite the observation of a recombinant between *FRAXA* and *DXS297* in a definite carrier female. This result presumably arises from the incorporation of mutation into the analysis: there is a small chance that the apparent "recombinant" is instead a second mutation in a family already segregating for the fragile

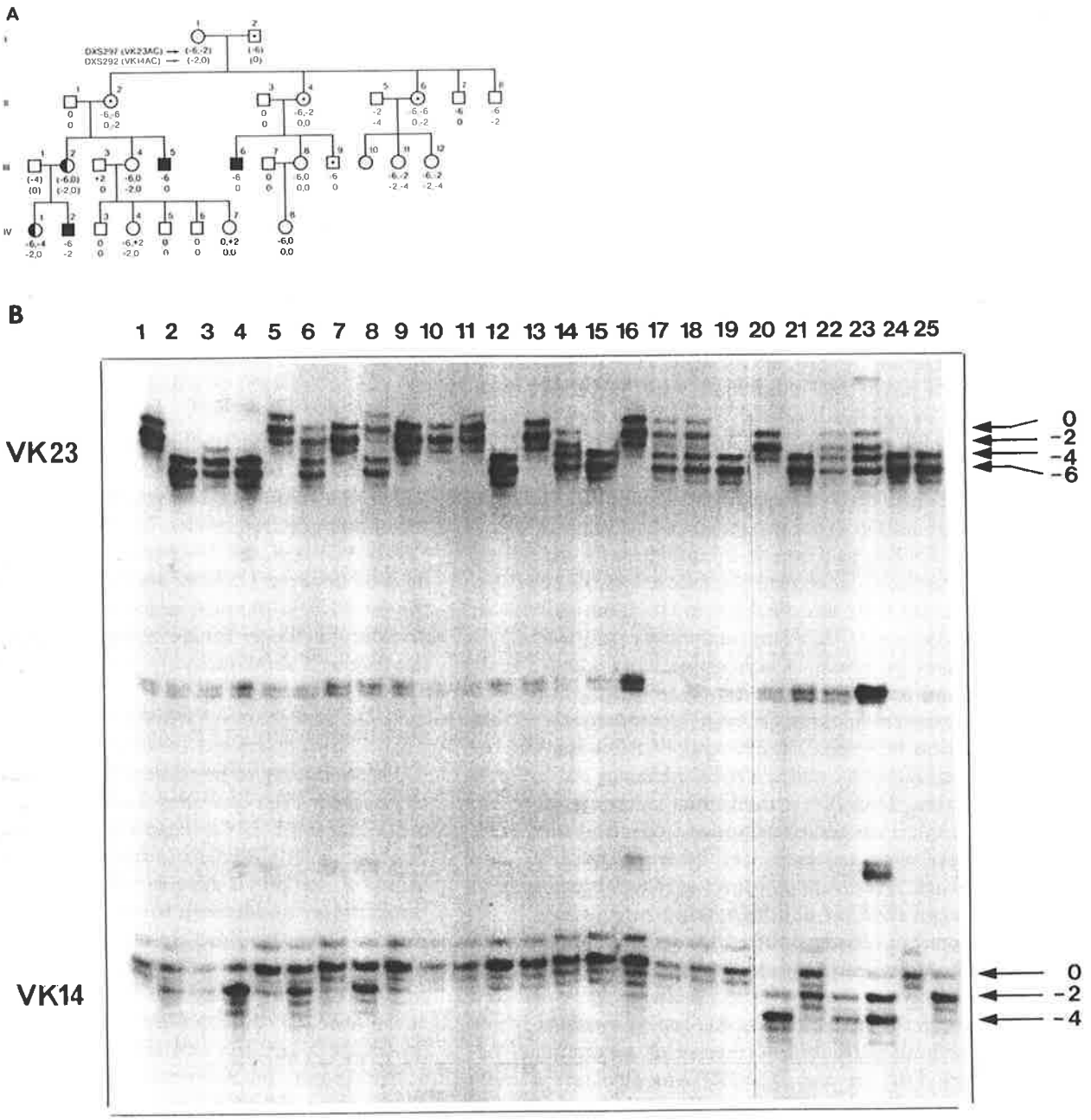



**Figure 2** Genotype analysis of VK14AC and VK23AC repeats in a fragile X syndrome pedigree. *Right,* Fragile X syndrome pedigree with DXS297 (AC) and DXS292 (AC) genotypes. ⊡ and ⊙ = transmitter based on pedigree and/or DNA results; ◐ and ■ = affected individual with cytogenetically characterized fragile X; ◑ = affected individual without cytogenetic analysis; ( ) = genotype inferred. Carrier status was determined in conjunction with analysis of the distal flanking marker St14. *Below,* Autoradiograph of DXS297 (AC) and DXS292 (AC) genotypes. Individuals in the affected pedigree are as follows: lane 1, II-1; lane 2, II-2; lane 3, IV-1; lane 4, IV-2; lane 5, III-3; lane 6, III-4; lane 7, IV-3; lane 8, IV-4; lane 9, IV-5; lane 10, IV-6; lane 11, IV-7; lane 12, III-5; lane 13, II-3; lane 14, II-4; lane 15, III-6; lane 16, III-7; lane 17, III-8; lane 18, IV-8; lane 19, III-9; lane 20, II-5; lane 21, II-6; lane 22, III-11; lane 23, III-12; lane 24, II-7; and lane 25, II-8.

## References


Chamberlain JS, Gibbs RA, Ranier JE, Nguyen PN, Caskey CT (1988) Deletion screening of the Duchenne muscular dystrophy locus via multiplex DNA amplification. Nucleic Acids Res 16:11141–11156

Conneally PM, Edwards JH, Kidd KK, Lalouel J-M, Morton NE, Ott J, White R (1985) Report of the Committee on Methods of Linkage analysis and Reporting. Cytogenet Cell Genet 40:356–359

Daussett J, Cann H, Cohen D, Lathrop M, Lalouel J-M, White R (1990) Centre d'Etude du Polymorphisme Humain (CEPH): collaborative genetic mapping of the human genome. Genomics 6:575–577

Hofker MH, Skraastad MI, Bergen AAB, Wapenaar MC, Bakker E, Millington-Ward A, van Ommen GJB, Pearson PL (1986) The X chromosome shows less genetic variation at restriction sites than the autosomes. Am J Hum Genet 39:438–451

Hyland VJ, Fernandez KEW, Callen DF, MacKinnon RN, Baker E, Friend K, Sutherland GR (1989) Assignment of anonymous DNA probes to specific intervals of human chromosomes 16 and X. Hum Genet 83:61–66

Kogan SC, Doherty BS, Gitschier J (1987) An improved method for prenatal diagnosis of genetic diseases by analysis of amplified DNA sequences. N Engl J Med 317:985–990

Litt M, Luty JA (1989) A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. Am J Hum Genet 44:397–401

Mulley JC, Turner G, Bain S, Sutherland GR (1988) Linkage between the fragile X and *F9, DXS52* (St14), *DXS98* (4D-8) and *DXS105* (cX55.7). Am J Med Genet 30:567–580

Oberlé I, Heilig R, Moisan JP, Kloepfer C, Mattei MG, Mattei JF, Boue J, et al (1986) Genetic analysis of the fragile X mental retardation syndrome with two flanking polymorphic DNA markers. Proc Natl Acad Sci USA 83: 1016–1020

Richards RI, Holman K, Shen Y, Kozman H, Harley H, Brook D, Shaw D. Human glandular kallikrein genes: genetic and physical mapping of the KLK1 locus using a highly polymorphic microsatellite PCR marker (submitted)

Sutherland GR, Baker E (1990) The common fragile site in band q27 of the human X chromosome is not coincident with the fragile X. Clin Genet 37:167–172

Sutherland GR, Hecht F (1985) Fragile sites on human chromosomes. Oxford University Press, Oxford

Sutherland GR, Mulley JC (1990) Diagnostic molecular genetics of the fragile X. Clin Genet 37:2–11

Suthers GK, Hyland VJ, Callen DF, Oberle I, Rocchi M, Thomas NS, Morris CP, et al (1990) Physical mapping of new DNA probes near the fragile X (*FRAXA*) by using a panel of cell lines. Am J Hum Genet 47:187–195

Suthers GK, Mulley JC, Voelckel MA, Dahl H, Väisänen ML, Steinbach P, Glass IA, et al (1991*a*) Genetic mapping of new DNA probes at Xq27 defines a strategy for DNA studies in the fragile X syndrome. Am J Hum Genet 48: 460–467

Suthers GK, Oberlé I, Nancarrow J, Mulley JC, Hyland VJ, Wilson PJ, McCure J, Morris CP, Hopwood JJ, Mandel JL, Sutherland GR (1991*b*) Genetic mapping of new RFLPs at Xq27-q28. Genomics 9:37–43

Weber JL, May PE (1989) Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am J Hum Genet 44:388–396

**Table 2**

**Two-Point Lod Scores for Fragile X Kindreds**

| Linkage Comparison | Lod Score at Recombination Fraction of | | | | | | | Maximum Recombination Fraction | Maximum Lod Score |
|---|---|---|---|---|---|---|---|---|---|
| | .0 | .01 | .05 | .1 | .2 | .3 | .4 | | |
| *FRAXA:DXS297*..... | 31.45 | 32.04 | 30.36 | 27.73 | 21.38 | 14.35 | 6.91 | .01 (.00–.04) | 32.04 |
| *FRAXA:DXS292*..... | ... | 9.95 | 12.70 | 12.72 | 10.67 | 7.51 | 3.81 | .07 (.03–.15) | 12.87 |
| *DXS297:DXS292*..... | ... | 7.82 | 11.57 | 12.06 | 10.48 | 7.52 | 3.85 | .09 (.04–.17) | 12.08 |

X. In contrast, the lod score for the *FRAXA:DXS292* comparison approaches minus infinity at a recombination fraction of zero, because the two recombinants involve unaffected individuals who have been diagnosed as carriers by closely linked flanking markers. Recombination events within the *DXS292–DXS297–FRAXA* linkage group were consistent with the established order derived from physical mapping. Only one recombination event was observed between *DXS297* and *FRAXA,* in a female unambiguously expressing the fragile X. Since recombination was observed between *DXS292* and *FRAXA* in the same individual, *DXS297* and *DXS292* are on the same side of *FRAXA*. In other pedigrees, a noncarrier female and a male transmitter (as determined by closely linked informative markers flanking *FRAXA*) demonstrated recombination between *DXS292* and *FRAXA* but not between *DXS297* and *FRAXA*. Hence, of the two markers, *DXS297* is confirmed as being closer to *FRAXA*. Two additional individuals, both carrier females, were recombinants between *DXS292* and *FRAXA*. They were uninformative for recombination between *DXS297* and *FRAXA*.

Figure 2 demonstrates Mendelian inheritance of both AC-repeat polymorphisms. Some of the difficulties in this large affected pedigree were resolved through the analysis of the AC-repeat markers, which were both partially informative. A recombinant between *FRAXA* and *DXS292* was detected between individuals II-2 and III-2. Male transmitter status for I-2 and III-9 was based on informative flanking markers VK23AC and St14, and III-8 and IV-8 are likely carriers.

The degree of linkage disequilibrium between VK23AC, the VK23B *Hin*dIII RFLP, and the VK23B *Xmn*I RFLP was not quantified. Only one family not informative for VK23AC was found to be informative for the *Hin*dIII RFLP, suggesting that the RFLP does not significantly increase informativeness in families already typed for VK23AC. Very few families have been typed for the *Xmn*I RFLP, so the extent to which this RFLP might increase informativeness in families already typed for VK23AC is not known.

### *Multipoint Linkage Analysis*

The result of multipoint analysis of VK23AC and VK14AC in CEPH families is given in figure 3. This shows the two-point lod scores and the multipoint recombination frequencies. The new marker *DXS292* (VK14AC) is placed 3% proximal to *DXS297* in the interval between *DXS297* and *DXS98*. This confirms the established physical order and is consistent with anecdotal evidence for the order derived above from the fragile X families.

## Discussion

The mutation responsible for fragile X syndrome has not yet been identified and characterized at the molecular level. DNA diagnosis relies on analysis of flanking genetic markers (Suthers et al. 1991*a*). The utility of markers is governed by their distance from the mutation and by their information content. In the present work, we have characterized two additional markers, both highly polymorphic and closely linked to *FRAXA*. The AC-repeat microsatellites identified at the *DXS297* and *DXS292* loci were applied to the 36 fragile X families available to our laboratory.

Multipoint recombination frequencies between markers near *FRAXA* were previously estimated from CEPH pedigrees (Suthers et al. 1991*b*); they are *DXS98* (12.3%) *DXS369* (0%) *DXS297* (5.7%) *DXS296* (0%) *IDS* (1%) *DXS304* (12%) *DXS52* (recombination frequency is shown in parentheses between the respective markers). *FRAXA* was located 3.7% distal to *DXS297* and 2% proximal to *DXS296* (Suthers et al. 1991*a*). In the present study the addition of *DXS297* (VK23AC) and *DXS292* (VK14AC) data from the CEPH families had little effect on the multipoint recombination frequencies between the

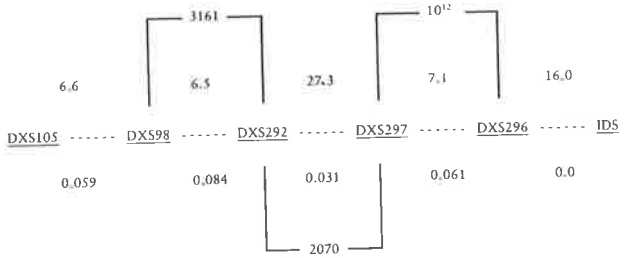



**Figure 3** Partial multipoint map of region near *FRAXA*, based on CEPH analysis. *FRAXA* is located between *DXS297* and *DXS296*. The order of loci was derived from physical mapping and confirmed by the odds, as shown, against inverting adjacent loci. Numbers above the marker intervals are two-point lod scores, and numbers below the marker intervals are multipoint recombination fractions.

markers (fig. 2). The suggested recombination frequency for diagnosis using VK23AC therefore remains at 4%; however, when considered in light of subsequent data from the fragile X families themselves, as determined in the present study, this may be an overestimate (table 2). The previous report of 4% was based only on the VK23B *Hin*dIII RFLP typed in a subset of the Adelaide families. The data presented in table 2 were based on all the Adelaide families and used the more informative VK23AC marker. When data are available from additional markers currently being characterized near the fragile X, the *FRAXA* locus may be repositioned onto the background map, as previously described by Suthers et al. (1991*a*).

*DXS292* has now been placed on the multipoint background map by using CEPH pedigrees. Since it is 3 cM proximal to *DXS297* on the background map, the suggested recombination frequency for diagnosis of fragile X syndrome is 7%. This corresponds exactly with the most likely recombination frequency derived from fragile X families by using two-point analysis (table 2). When it is used for diagnosis, *DXS292* should be used in conjunction with an informative marker distal to *FRAXA*.

The availability of numerous closely linked markers (Suthers et al. 1991*a*) can have considerable impact on the precision of linkage analysis. In the present study, sets of closely linked informative markers flanking *FRAXA* facilitated the identification with virtual certainty of male transmitters and carrier females. Although it has been shown that recombination values are insensitive to variation of the penetrance parameters used in linkage analysis (Oberlé et al. 1986), the exact coding of carrier status, wherever possible, had substantial impact on both the magnitude of lod scores and the associated lod – 1 confidence intervals.

Three mentally retarded individuals (not expressing the fragile X) within our set of fragile X families were shown (with a probability greater than 99%) not to have inherited the chromosomal segment containing *FRAXA*. This excluded the diagnosis of fragile X–linked mental retardation. The diagnosis of five nonretarded individuals expressing the "fragile X" in 1%–2% of cells was also clarified. None of these individuals had inherited the chromosomal segment containing the fragile X, implying that the detection of the common fragile site (Sutherland and Baker 1990) can lead to misdiagnosis of carrier status. In two cases this confirmed the conclusion tentatively made earlier when only loosely linked flanking markers were available (Mulley et al. 1988). The availability of numerous flanking markers can be invaluable for accurate diagnosis of some mentally retarded individuals and of individuals with low rates of expression of the "fragile X." These clarifications can have considerable impact on the determination of the potential carrier status of other family members in some sections of a pedigree.

Finally, the simultaneous detection of highly polymorphic VK23AC and VK14AC markers as described above now allows determination of at least one informative marker on the proximal side of *FRAXA*—within 3 d, in most families. Because these markers are highly polymorphic, they facilitate detection of both nonpaternity and sample error and often permit inference of missing parental marker genotypes when this cannot be achieved with the less informative diallelic RFLPs. The present experimental approach's failure to secure polymorphic markers from VK18 and VK37 necessitates either screening additional sequences distal to *FRAXA* as they become available or walking from the existing loci (e.g., with yeast artificial chromosomes), in order to identify polymorphic dinucleotide repeats. The characterization of these markers distal to *FRAXA* will enable a comprehensive approach to genetic linkage, an approach that will supersede present RFLP analysis.

## Acknowledgments

R.I.R. wishes to thank Shelley Richards for support and encouragement during the course of these studies. This work was supported in part by grants from the Channel 7 Children's Research Foundation of South Australia, Incorporated, and from the National Health and Medical Research Council of Australia. Y.S. was supported by a fellowship from the World Health Organization.

# Dinucleotide repeat polymorphism at the D16S288 locus

Y.Shen, K.Holman, A.Thompson, H.Kozman, D.F.Callen, G.R.Sutherland and R.I.Richards
Department of Cytogenetics and Molecular Genetics, Adelaide Children's Hospital, North Adelaide. SA 5006. Australia

*Source and Description of Clone*: A cosmid library (1) of flow-sorted human chromosome 16 from the somatic hybrid cell line CY18 was digested with *Sau*3AI. subcloned into *Sal*I cut M13 mp 19 after partial end-filling of both restriction sites and screened by hybridization to poly(dC-dA)·poly(dG-dT). One of the positive cloned fragments was sequenced and designated 16A-C7.1. The length polymorphism of the AC repeat contained in this subclone was typed using the polymerase chain reaction (PCR). The length of the amplified fragment (A3 allele) was 160 bp.

*PCR Primers*:
ATGAGGAGCAGAGGTGTCTCTT (CA strand)
TAAATCACAGACTCCCTTGGCT (GT strand)

*Frequencies*: Estimated from 146 chromosomes of unrelated CEPH parents.
Het = 0.73 PIC = 0.69

| Allele (bp) | Frequencies | Allele (bp) | Frequencies |
|---|---|---|---|
| A6 (154) | 0.034 | A3 (160) | 0.226 |
| A5 (156) | 0.158 | A2 (162) | 0.411 |
| A4 (158) | 0.164 | A1 (166) | 0.007 |

*Chromosomal Localization*: Assigned to chromosome 16p11.2−12.1 by PCR using DNA templates isolated from a panel of somatic cell hybrids (2).

*Mendelian Inheritance*: Co-dominant segregation was observed in CEPH families 45 and 1334.

*PCR Conditions and Other Comments*: The PCR reaction was performed on 100 ng of genomic DNA with 150 ng of each oligodeoxyribonucleotide primer and 10 $\mu$Ci alpha $^{32P}$dCTP using a previously described reaction mixture (3). Conditions of the PCR amplification are as follows: 10 cycles of 1 minute at 94°C. 1.5 minutes at 60°C. and 1.5 minutes at 72°C. 25 cycles of 1 minute at 94°C. 1.5 minutes at 55°C. and 1.5 minutes at 72°C. The final elongation cycle was 10 minutes at 72°C. The amplified alleles were resolved on denaturing polyacrylamide gels and detected by autoradiography. The dinucleotide repeat was based on a AACAGG$(GT)_{22}$ACATCT sequence.

*Acknowledgements*: This work was supported by grants from the United States Department of Energy (Contract DE-FG02−89ER60863) and the National Health and Medical Research Council of Australia. We thank Drs. Ray Stallings and Ed Hildebrand for an aliquot of the human chromosome 16 cosmid library.

*References*: 1) Stallings,R.L. *et al.* (1990) *Proc. Natl. Acad. Sci. USA* **87**, 6218−6222. 2) Callen,D.F. *et al.* (1990) *Ann. Génét.* **33**, 190−195. 3) Richards,R.I. *et al.* (1991) *Am. J. Hum. Genet.* **48**, 1051−1057.

# MboI RFLP at the D4S43 (C4H) locus

L.P.Lazarou, R.G.Snell, M.E.MacDonald[1], J.F.Gusella[1], J.J.Wasmuth[2] and D.J.Shaw
Institute of Medical Genetics, University of Wales College of Medicine, Heath Park, Cardiff CF4 4XW, UK, [1]Molecular Neurogenetics Laboratory, Massachusetts General Hospital, Harvard University, Boston, MA 02114 and [2]Department of Biological Chemistry, D240 Med. Sci. I, University of California, Irvine, CA 92717, USA

*Source/Description*: Probe PKP 1.65. 1.65 kb PstI fragment subcloned into pGEM 3 (2.7 kb) (MacDonald *et al.*, 1989).

*Polymorphism*: Probe PKP 1.65 identifies a two allele polymorphism with the restriction endonuclease MBO I.

*Frequency*: Studied in 141 normal chromosomes

| | | |
|---|---|---|
| MBO I | (L1) 1.25 kb | 0.43 |
| | (L2) 0.66 kb | 0.57 |

*Chromosomal Localisation*: The probe is found within the D4S43 (C4H) locus and maps to 4q16.3 (Gilliam *et al.*, 1987). The polymorphism is contained within hybrids JS4 A9.1 (4 only) and HHW 693 (pter-15.1).

*Mendelian Inheritance*: Co-dominant inheritance shown in 68 meioses.

*Probe Availability*: Available upon request from M.E.MacDonald.

*Other Comments*: The polymorphism will increase the informativeness of the D4S43 locus in HD presymptomatic testing.

*References*: 1) Gilliam *et al.* (1987) *Science* **238**, 950−952. 2) MacDonald *et al.* (1989) *J. Clin. Inv.* **84**, 1013−1016.

# An STS at the D16S290 locus

**L.Z.Chen, Y.Shen, K.Holman, A.Thompson, S.Lane, R.I.Richards, G.R.Sutherland and D.F.Callen**

Department of Cytogenetics and Molecular Genetics, Adelaide Children's Hospital, North Adelaide, South Australia 5006, Australia

Submitted August 13, 1991

A DNA clone (T102) prepared from a cosmid library of flow-sorted human chromosome 16 (1) screened by hybridization to a tetranucleotide repeat $(AGAT)_n$ was sequenced. T102 contains a poly (T) tract of 40 nucleotides interrupted by five cytosines. Using PCR conditions described below, a fragment of the expected size (275 bp) was amplified and located at 16q12.1 by analysis of the mouse-human somatic cell hybrid-panel (2, 3) in the interval between CY132 and CY140 (see figure). PCR analysis of 20 unrelated caucasian individuals failed to detect any polymorphism at this locus.

*PCR Primers*:
Forward CGTAAGAAGAGGTTTGCACC
Reverse CAAGAGCGAAACTCCGTCTC

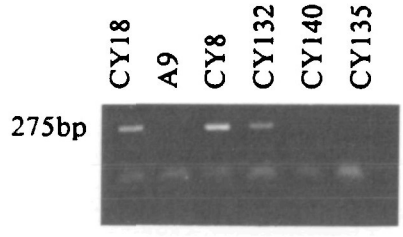
*PCR Components*:
100 ng of human genomic DNA,
150 ng of each oligodeoxyribonucleotide primer,
0.5 U Amplitaq DNA Polymerase (Perkin Elmer Cetus), and
7 mM $MgCl_2$ in 20 $\mu$l of PCR reaction mix (4).

*PCR Profile*:
94°C for 1 minute
60°C for 1.5 minutes
72°C for 1.5 minutes for 10 cycles
94°C for 1 minute
55°C for 1.5 minutes
72°C for 1.5 minutes for 25 cycles
The final elongation cycle was 10 minutes at 72°C.

*Sequence of the PCR product*:
CGTAAGAAGAGGTTTGCACCCCACCTGGCC-
TGGCCGTGTCCTCTGGGCTGCCCTGTCGGCTGCCTGC-
TGGGCTTAGCTTAAGAATGGCTGAGATGAATCTGTGC-
TGGGCAGGGGACAGTCCTGCTGCATCTGAAAGAAGG-
GTGCCCCATATCCCACTCATTGGGGATGGTGGCCTG-
GGCTCTAGGGTCGCCATGTGGGAGATGAGATGGATT-
TTTTTTTCTTTTCTTTTTTCTTTTTCTTTTTCTTTTTTT-
TTTTGAGACGGAGTTTCGCTCTTG

## REFERENCES

1. Stallings,R.L. *et al.* (1990) *Proc. Natl. Acad. Sci. USA* **87**, 6218–6222.
2. Callen,D.F. *et al.* (1990) *Ann. Genet.* **33**, 190–195.
3. Chen,L.Z. *et al.* (1991) *Genomics* **10**, 308–312.
4. Richards,R.I. *et al.* (1991) *Am. J. Hum. Genet.* **48**, 1051–1057.



GENOMICS 13, 402–408 (1992)

# Isolation and Characterisation of $(AC)_n$ Microsatellite Genetic Markers from Human Chromosome 16

A. D. THOMPSON, Y. SHEN, K. HOLMAN, G. R. SUTHERLAND, D. F. CALLEN, AND R. I. RICHARDS

*Department of Cytogenetics and Molecular Genetics, Adelaide Children's Hospital, North Adelaide SA 5006, Australia*

Received October 30, 1991; revised February 4, 1992

**A cosmid library of human chromosome 16 has been subcloned, and $(AC)_n$ microsatellite positive clones have been identified and sequenced. Oligonucleotide primers flanking the repeat were designed and synthesized for $(AC)_n$ microsatellites with $n > 16$. These microsatellite loci were then mapped by PCR using a somatic cell hybrid panel of human chromosome 16, and their heterozygosities and allele frequencies determined. Fourteen $(AC)_n$ microsatellites were mapped to discrete physical intervals of human chromosome 16 defined by a mouse/human hybrid panel. Nine of these have expected heterozygosities ranging between 0.60 and 0.79, four have expected heterozygosities between 0.02 and 0.49, and one detected three loci where the alleles could not be resolved.**



## INTRODUCTION

$(AC)_n$ microsatellite loci are distributed along the length of human chromosome 16 at a frequency of approximately one locus every 30 kb (Stallings *et al.*, 1991). The number of dinucleotide (AC) pairs at these loci has been shown to be polymorphic (Litt and Luty, 1989; Weber and May, 1989; Luty *et al.*, 1990). They comprise either uninterrupted or interrupted runs of consecutive (AC) dinucleotide pairs. Other simple repeats may be present in addition to the $(AC)_n$ repeat. Weber (1990) has assigned these repeats to three classes, perfect, imperfect, and compound respectively, and shown that their PIC value is proportional to the length of the perfect repeat (Weber, 1990).

These microsatellites are therefore an abundant source of highly informative PCR-based genetic markers that appear to be evenly distributed throughout the human genome. Traditional RFLPs with high PIC values tend to be VNTRs that are usually clustered near the telomeres. For chromosome 16, only seven single-enzyme RFLPs with PIC values of greater than 0.50 have been reported (Reeders and Hildebrand, 1989; Reeders *et al.*, 1990), but of four $(AC)_n$ microsatellite probes mapped to chromosome 16, three had PIC values greater than 0.50, and the fourth had a PIC value of 0.44 (Reeders *et al.*, 1990).

$(AC)_n$ microsatellites are ideal polymorphic loci for the construction of high-resolution genetic maps of human chromosomes. They also provide a means of making candidate disease gene loci informative, allowing segregation analysis in affected pedigrees (Richards *et al.*, 1991a,b,c). To meet these two complementary requirements $(AC)_n$ microsatellites have been isolated by two approaches: either by sequence analysis of previously characterized genes and anonymous DNA sequences or by isolation at random from human DNA libraries. In this report the latter approach has been applied to chromosome 16.

## MATERIALS AND METHODS

*Subcloning of human chromosome 16 for sequencing.* Total DNA was prepared from a cosmid library of human chromosome 16 (Stallings *et al.*, 1990). This library was constructed from flow-sorted chromosomes from the mouse–human somatic cell hybrid CY18, which has chromosome 16 as its only human chromosome (Callen, 1986). The DNA was digested to completion with *Sau*3AI (New England Biolabs) and ligated into the *Bam*HI site of M13mp18. A preliminary experiment used total digested DNA and a range of insert-to-vector ratios. Subsequently, the *Sau*3AI-digested DNA was size-fractionated on a 1.4% low-melting-point agarose gel, and the size range 200–500 bp ligated with the M13mp18–*Bam*HI vector at a threefold molar excess of vector.

*Selection of $(AC)_n$ positive clones.* M13mp18 cosmid library ligated DNA was transformed into *Escherichia coli* strain BB4 (Bullock *et al.*, 1987), and plaque lifts were made using Hybond $N^+$ gridded nylon filters (Amersham). Filters were hybridized in the absence of carrier DNA at 65°C in 1 *M* sodium phosphate (pH 7.0), 7% SDS and probed with poly$(AC)_n$/$(GT)_n$ (Pharmacia) labeled with [$\alpha$-$^{32}$P]dCTP (Amersham) using a multiprime labeling kit (Amersham). Washes were performed to a maximum stringency of 30 m*M* NaCl, 2.3 m*M* $NaH_2PO_4$, 0.25 m*M* EDTA, 0.1% (v/v) SDS (adjusted to pH 7.4 with NaOH).

From a preliminary experiment, single-stranded DNA was prepared from all plaques showing strong hybridization to the probe and sequenced. In the subsequent experiment, all positive plaques were spotted onto a fresh bacterial lawn and screened a second time as described above. Single-stranded DNA prepared from these second-screen positives was spotted onto GeneScreen Plus (DuPont). The dot-blot filter was screened with the poly$(AC)_n$/$(GT)_n$ probe, and those clones giving the strongest signals were selected for sequencing.

*Oligodeoxyribonucleotides.* Initially, primers were designed for $(AC)_n$ repeats when $n$ was >16 for both perfect and imperfect repeats, but later only for perfect repeats with $n > 16$. Primers were usually 25 bases in length and as close to 50% GC content as the flanking sequence allowed. Primer pairs were designed to give PCR products



0888-7543/92 $5.00

ranging from 100 to 200 bp in length to facilitate the simultaneous PCR amplification of two or more loci (Weber and May, 1989).

*Panel of chromosome 16 somatic cell hybrid breakpoints.* The panel of somatic cell hybrids containing breakpoints of human chromosome 16 has been documented elsewhere (Callen, 1986; Callen *et al.*, 1990; Richards *et al.*, 1991d; Callen *et al.*, unpublished results). The positions of hybrid breakpoints have, on many occasions, been determined by negative results from Southern blot analysis. In some instances PCR analysis has given results inconsistent with the Southern blot data. For CY123 and CY140, further Southern blot analysis confirmed the presence of more of chromosome 16 than the initial results suggested, and their breakpoint locations have been altered accordingly (Callen *et al.*, unpublished results).

*PCR Mapping and characterization of $(AC)_n$ loci.* PCR was performed on a Perkin–Elmer/Cetus DNA thermal cycler using the reaction conditions of Kogan *et al.* (1988) with the addition of 5 $\mu$Ci of [$\alpha$-$^{32}$P]dCTP for polymorphism typing. The 10-$\mu$l reaction mixture contained 16.6 m*M* ammonium sulfate; 67 m*M* Tris–HCl (pH 8.8); 10 m*M* $\beta$-mercaptoethanol; 6.7 $\mu$*M* EDTA; 170 $\mu$g/ml bovine serum albumin; 100 ng genomic DNA, 1.5 m*M* each dATP, dCTP, dGTP, and dTTP; 75 ng each oligodeoxyribonucleotide primer; 10% (v/v) dimethyl sulfoxide; 5 $\mu$Ci [$\alpha$-$^{32}$P]dCTP (3000 Ci/mmol, Amersham); and 0.5 unit *Taq* polymerase (Perkin–Elmer/Cetus). The magnesium chloride concentration was optimized for each marker by analysis of the above reaction without [$\alpha$-$^{32}$P]dCTP on 1.4 or 2% agarose gels. The optimum magnesium chloride concentration for each marker is given in Table 1. All mapping and polymorphism typing reactions employed a step cycle file comprising 10 cycles of 94°C × 1 min, 60°C × 1.5 min, 72°C × 1.5 min, followed by 25 cycles of 94°C × 1 min, 55°C × 1.5 min, 72°C × 1.5 min, and a 10-min extension at 72°C.

The physical location of each marker was determined by PCR analysis using DNAs from a human chromosome 16 somatic cell hybrid panel (Callen *et al.*, 1990; Richards *et al.*, 1991d; Callen *et al.*, unpublished results) as templates. PCR products were resolved on 1.4 or 2.0% agarose gels and visualized by ethidium bromide staining. This allowed mapping of the marker to a specific breakpoint interval by virtue of the presence or the absence of a PCR product with somatic cell hybrid DNAs as templates in the PCR. PCR for hybrid DNAs with breakpoints that flanked this interval helped to confirm the physical assignment.

Heterozygosities and allele frequencies were determined by PCR of DNAs from unrelated Caucasian individuals and/or CEPH (Centre d'Etude du Polymorphisme Humain, Paris) parent DNAs. PCR products were resolved on 5% denaturing polyacrylamide gels following the addition of a threefold excess of formamide loading buffer and heating at 94°C for 3 min. Electrophoresis of PCR products on denaturing polyacrylamide gels was essential to distinguish alleles that frequently differ by as little as 2 bp in length.

## RESULTS

### *Subcloning and Sequencing of $(AC)_n$ Positive Clones*

Of the 48 clones sequenced, 44 were unique and 4 were independently subcloned from the chromosome 16 cosmid library twice. Oligodeoxyribonucleotides were designed for 23 $(AC)_n$ repeats and are given in Table 1. Of the remainder, 2 contained no discernible repeat and 11 were rejected as too short, having perfect repeats of $n < 16$. Three had repeats too close to the cloning site to design primers, 2 were flanked by repetitive sequence, 2 needed very long sequencing runs, and the last was a mixed or unstable clone.

### *Physical Mapping of $(AC)_n$ Markers*

Physical mapping proceeded in two stages. First, DNAs from a set of hybrids comprising CY2, CY4, CY8, CY115, CY165, and CY196 were used as templates, allowing localization to a broad interval. A second round of PCR was then performed with each of the hybrid breakpoint DNAs that either localized within or flanked the initial broad interval. In this manner the physical location of a marker was defined by a minimum of two hybrid breakpoints on either side of the interval to which it mapped.

From the 23 $(AC)_n$ microsatellite loci for which primers were designed, 14 were physically mapped to discrete intervals of the human chromosome 16 by PCR analysis of DNAs from a somatic cell hybrid panel of chromosome 16 (Fig. 1). PCR amplification of CY18 DNA (which contains a single intact human chromosome 16 and no other human chromosomes; Callen *et al.*, 1990) showed that two of these primer pairs detected more than one locus. In both cases each locus was within the same physical interval defined by the hybrid panel. 16AC5.4 detected two loci whose PCR products resolved well and could be typed independently, but 16AC6.16 detected three loci whose PCR products had similar mobilities. Since specific alleles could not be assigned to specific loci, 16AC6.16 could not be used as a genetic marker.

16AC1.8 mapped to human chromosome 16, but generated a complex pattern when used against the hybrid panel. Both CY18 DNA and total human DNA gave a ladder of PCR products, indicating that 16AC1.8 detects a number of loci on chromosome 16.

Four of the remaining eight unmapped $(AC)_n$ microsatellite primer pairs, AC2, 1.16, 6.18, and 6.24, gave discrete bands with mouse control DNA but not with human DNA. Clones 1.17 and 2.1 gave multiple bands on background smears for both human and mouse genomic DNA. 6.10 gave faint multiple bands on a background smear for human and CY18 DNA, but not mouse, and 6.7 gave inconsistent results with human and CY18 DNA, but no products with mouse DNA.

Allele frequencies and heterozygosities were determined (Table 2) for all $(AC)_n$ microsatellite loci mapped to discrete intervals of the human chromosome 16 hybrid panel except for the three loci detected by 16AC6.16. Primers for two different $(AC)_n$ repeats were often combined in a multiplex reaction, thereby increasing the efficiency of typing multiple individuals.

## DISCUSSION

Thirteen human $(AC)_n$ microsatellites have been isolated from a cosmid library of human chromosome 16 and characterized with regard to their allele frequencies and heterozygosities. Nine of these represented perfect $(AC)_n$ repeats of $n > 18$ and heterozygosities of >0.60. The remaining four had perfect $(AC)_n$ repeats of $n < 11$, but imperfect repeats of $24 < n < 32$, and heterozygosities ranging from 0.02 to 0.49. These four repeats were selected for primer synthesis early in this study, and

## TABLE 1

**Oligodeoxyribonucleotide Primer Sequences Designed for Selected $(AC)_n$ Repeat Loci**

| Clone | Locus | GT strand primer | $(GT)_n$ repeat sequence | AC strand primer | PCR product size in bp[a] ($Mg^{2+}$ optimum, m$M$)[b] |
|---|---|---|---|---|---|
| 2.5 | 16AC2.5 D16S291 | GCAGCCTCAGTTGTGTTTCCTAATC | $GGG(GT)_{25}GAT$ | AGTGCTGGGATTACAGGCATGAACC | 170 (4.0) |
| 2.3 | 16AC2.3 D16S292 | GGCATGTCAGGCCAGCCATGTTTT | $TAT(GCT)_7(GT)_{18}TTC$ | CTTTGCACAAAAACAGTAGCTATCCAC | 180 (3.0) |
| 6.16 | 16AC6.16 D16S293 | TGTCCCTGCTGAGCCTACTGTGCA | $GCA(GT)_6AT(GT)_{18}TTG$ | AATTCTGGCCACTACCACCCTGGTA | 141 (4.5) |
| AC1 | 16AC1 D16S294 | CACGGTTCCAAGCACGGCAAGAGAG | $CAG(GT)_7AT(GT)_4GCGTGC(GT)_6$-$GCGTGCGTGC(GT)_6GAA$ | GACCCGGGTAACAGAAAGAGACTC | 138 (4.5) |
| 3.12 | 16AC3.12 D16S298 | TTCCTCATGTATAAATTGGGTGTGGCCA | $GTA(GT)_8GC(GT)_{24}TAA$ | TGGTCCCGGCCCAATCCCAATGCTT | 188 (4.0) |
| 6.17 | 16AC6.17 D16S299 | TCCAACTGCTGGGATTACAGGCACA | $GGG(GT)_{19}TTG$ | ACAGAGTGAGGACCCCATCTCTATC | 124 (4.0) |
| 1.1 | 16AC1.1 D16S300 | AGCCAAGCTAGTTAATTTGTGGTCCCA | $GTT(GA)_{12}(GT)_{24}(GA)_5GCA$-$(GT)_6ATGGCA(GT)_6ATA$ | AAGCGCCTGGGAGAGAGCAAGCTAT | 171 (3.5) |
| 6.5 | 16AC6.5 D16S302 | AGACTTCCAGCATGGTCTCATTATC | $TTT(GT)_7TT(GT)_{10}AT(GT)_6GAG$ | GATCTAGAAGTCTCAGCATCTGGAA | 111 (5.0) |
| 1.14 | 16AC1.14 D16S304 | GTCAGTGCAATGGAGGTAAGAAAAAG | $ATC(GT)_{20}ATT$ | GATCAGATGAGATAGGGCATATTCATGG | 140 (4.5) |
| 1.18 | 16AC1.18 D16S308 | CAGCCAGGGTAGTAAGGCTAGACCT | $TTT(GT)_{27}GAG$ | TGGGTGGCAGAGTGAGACCCTGTCT | 163 (3.5) |
| 5.4 | 16AC5.4 D16S306 | TAAGGATTGAAGTCCCACTAGCCAG | $GGA(GT)_9ATGA(GT)_3GGCTCTTT$-$(GT)_3GG(GT)_2CT(GT)_4CTCT(GT)_6$-$CTGTAT(GT)_2TTCTCT(GT)_5CT$-$(GT)_7TTGTCT(GT)_3CTGCAT$-$(GT)_4ACT$ | CATGATTACGCGCTGCCTCCAGTAC | 200 (3.5) |
| 6.21 | 16AC6.21 D16S301 | GATCCTAAGGACAAATGTAGATGCTCT | $GGG(GT)_{19}GCA$ | AGCCACTTCCCAGAACTTGGCTTCC | 146 (4.0) |
| 1.15 | 16AC1.15 D16S305 | CCTCCCAGGTTCAGGCAATTCTTCT | $TTT(GT)_{27}(AT)_6AGTGC(GT)_6ATA$ | TAGGCGACAGAGTGGGACTCCATTA | 188 (3.5) |
| 6.26 | 16AC6.26 D16S303 | GATCAGTGCTCGTTTTTTTTGGTTTGG | $TTG(GT)_9GC(GT)_2GCGCGTGC(GT)_2$-$GCGCGTGCGCGTGC(GT)_7TTT$ | CAACAAGAGCGAAACTCGGTCTCAA | 113 (4.5) |
| 1.8 | 16AC1.8 | ACCGTGCCAGCCTCATGTCATTCTT | $CTT(GT)_{23}GAC$ | ATCACACCACTGTACTCCAGCCTGA | 113 |
| AC2 | Mouse | CTGAGCATGTTGGTCATTACTATGTC | $GTT(GT)_7CT(GT)_8ATT$ | CACATCCCACAAAACTGAAATCGTGG | 183 |
| 1.16 | Mouse | AGGTCACAGGCTAGGATATTAGCC | $GTG(GT)_{23}CT(GT)_6CT(GT)_2CT$-$(GT)_2CT(GT)_4CT(GT)_3CTG$ | CAAAGGCTGCATCACTGCTTCAGA | 180 |
| 6.18 | Mouse | TTAACCAGCTGAGCTTGAGCAAAGG | $ATA(GT)_{17}ACG$ | CACCTTCACCATGGCATATGTACGT | 118 |
| 6.24 | Mouse | CCAACTGTGGCCTAACTCATGCTCT | $CTT(GT)_3GA(GT)_3GAGA(GT)_{16}$-$(GA)_{16}(GGGA)_{11}GAA$ | AATCCTGAGGATGATACCCTCTGGT | 181 |
| 2.1 | —[c] | ACTGGGATTACAGGCATAAGCCACC | $TAT(GT)_{20}ATG$ | GGCAACAAGAGCAAAACTCTGTCTC | 174 |
| 1.17 | — | GATCGCTCTACCTGGCCTGGAGG | $AGG(GT)_{23}N_{41}(GT)_7TTT$ | AATGCCAAAGGCAAGTCCTGCCCTA | 199 |
| 6.7 | — | ATAGGCACATGCTGCCATGCCTGGC | $GGC(GT)_{21}ATT$ | GAACTCAGGAGTTTGCAACCAGCCT | 126 |
| 6.10 | — | TTCTCATGCCTCAGCCTCCTGAGTA | $AAT(GT)_{20}AT(GT)_2T(GT)_6ATT$ | ATCAGCCTGGGCAATGTGGCAAAAC | 165 |

[a] PCR product size for allele sequenced.
[b] $Mg^{2+}$ optimum determined only for those PCRs giving clear results in human genomic DNA.
[c] Failed to give satisfactory results with either mouse or human genomic DNA.

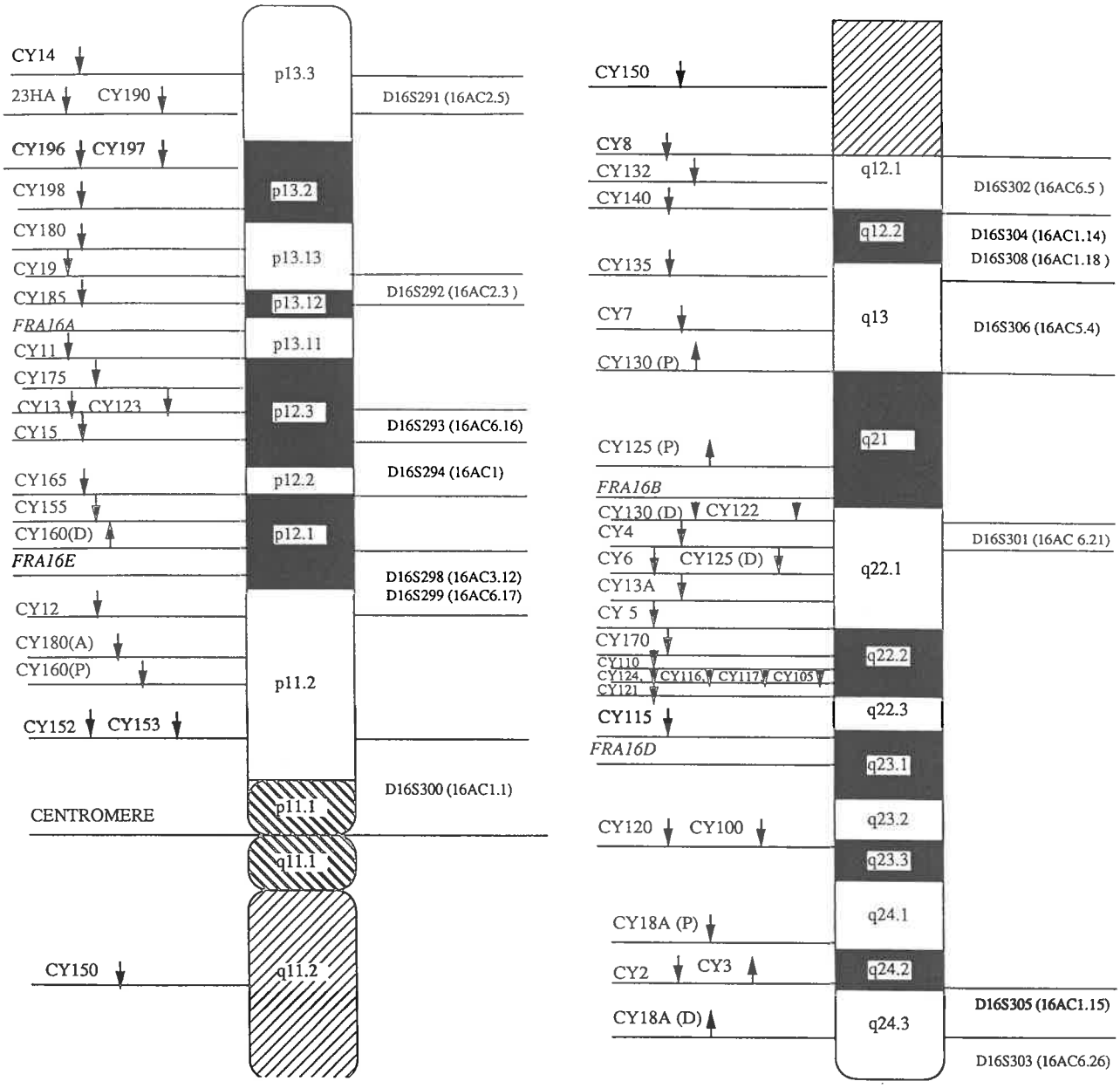



**FIG. 1.** Idiogram of chromosome 16 showing the distribution of (AC)$_n$ loci. Translocation and deletion breakpoints contained in the hybrid panel are shown on the **left** denoted by CY and a number. Fragile sites are given as *FRA16* and a letter. Arrows indicate the direction of DNA (from the breakpoint indicated to the telomere) contained in the hybrid, unless indicated by (P) or (D), which indicate the extent of proximal and distal boundaries, respectively. In the direction of the long arm the tip of the arrow touches the breakpoint, while in the direction of the short arm the base of the arrow touches the breakpoint. The positions of the breakpoints in the hybrids CY123 and CY140 have been reassessed since their previous publication (Richards *et al.*, 1991d). In both cases more of chromosome 16 was found to be present in the hybrid than previously thought.

after their low heterozygosities were observed only perfect repeats of $n > 16$ were selected subsequently. These results concur with the findings of Weber (1990) that the length of the perfect repeat is a better indicator of informativeness than the length of the imperfect repeat.

Three of the 15 PCR primer pairs amplifying human chromosome 16 (AC)$_n$ repeats detected three or more loci per individual. This is higher than the 7% cited by Weber (1990) and may reflect the small sample number reported here. Only one of the three, 16AC5.4, could be used as a genetic marker.

The large number of mouse (AC)$_n$ microsatellites isolated was an unexpected problem. Four of the 19 primer pairs that worked well represented mouse-derived (AC)$_n$ microsatellites. This is greater than the level of mouse DNA contamination of the cosmid library, which has been reported to be 7% (Stallings *et al.*, 1990). A possible explanation is that mouse (AC)$_n$ repeats appear on the average to be longer than human repeats, and therefore selection on the basis of length in the dot-blot screening step selected more strongly for mouse sequences than for human. It has also been reported that (AC)$_n$ microsatel-

**TABLE 2**—*Continued*

| Locus | Repeat | Alleles[b] | Frequencies | Expected heterozygosity | PIC value |
|---|---|---|---|---|---|
| 16AC1.18 (D16S308) | P(23) | 167 | 0.01 | 0.77 (92) | 0.73 |
| | | 165 | 0.02 | | |
| | | 163 | 0.04 | | |
| | | 161 | 0.18 | | |
| | | 159 | 0.25 | | |
| | | 157 | 0.36 | | |
| | | 155 | 0.05 | | |
| | | 153 | 0.02 | | |
| | | 149 | 0.03 | | |
| | | 137 | 0.02 | | |
| 16AC5.4 (D16S306) | P(9) + I(24) + I(19) max P(6) + max P(7) | 244 | 0.90 | 0.18 (124)[d] | 0.16 |
| | | 242 | 0.10 | | |
| 16AC6.21 (D16S301) | P(19) | 152 | 0.02 | 0.64 (154) | 0.58 |
| | | 150 | 0.06 | | |
| | | 148 | 0.08 | | |
| | | 146 | 0.47 | | |
| | | 144 | 0.36 | | |
| | | 142 | 0.01 | | |
| 16AC1.15 (D16S305) | CP(27) | 200 | 0.03 | 0.82 (108) | 0.80 |
| | | 198 | 0.04 | | |
| | | 196 | 0.01 | | |
| | | 194 | 0.05 | | |
| | | 192 | 0.06 | | |
| | | 190 | 0.01 | | |
| | | 188 | 0.05 | | |
| | | 186 | 0.12 | | |
| | | 184 | 0.07 | | |
| | | 182 | 0.11 | | |
| | | 180 | 0.37 | | |
| | | 178 | 0.01 | | |
| | | 176 | 0.05 | | |
| | | 174 | 0.01 | | |
| | | 172 | 0.01 | | |
| 16AC6.26 (D16S303) | I(32) max. P(9) | 115 | 0.07 | 0.43 (212) | 0.40 |
| | | 113 | 0.73 | | |
| | | 111 | 0.01 | | |
| | | 109 | 0.01 | | |
| | | 103 | 0.02 | | |
| | | 101 | 0.16 | | |

[a] P(25), perfect repeat of 25 repeat units; I, imperfect; c, compound.
[b] The size of the different alleles is given in base pairs.
[c] The number of chromosomes scored is given in parentheses.
[d] Although 16AC5.4 detects two loci, of 200 and 244 bp, only the larger one was polymorphic. The size of the larger band was estimated by comparison with the mobilities of sequences of known length in sequencing gels.

lites occur approximately twice as frequently in the mouse genome as in the human genome (Hamada *et al.*, 1982).

Other problems were that 17% of primer pairs failed to give satisfactory results with either human or mouse genomic DNA, and four clones were independently isolated and sequenced twice. Although a 17% primer failure rate is far from desirable, it is much lower than others have experienced (Cornall *et al.*, 1991). Nonrandom amplification of the cosmid library is most likely the reason that identical clones were independently isolated.

In conclusion, the strategy of randomly isolating (AC)$_n$ microsatellite loci from a human chromosome 16-specific library was successful in identifying highly informative and randomly distributed genetic markers specific to this chromosome. Fourteen (AC)$_n$ microsatellite loci have been mapped to 12 discrete intervals on the hybrid panel. Six of these loci have heterozygosities of >0.70, the cutoff proposed for index markers, and a further 3 were between 0.60 and 0.70. Gaps in the resulting genetic map could then be filled by employing a "targetted" approach, identifying and subcloning (AC)$_n$ repeats from lambda and cosmid clones previously mapped to the desired interval. This is possible only if detailed correlations exist between physical and genetic maps, as is the case with human chromosome 16 (Kozman *et al.*, unpublished results).

The characterization of AC repeat microsatellite markers therefore shows a great deal of promise for providing a high-resolution, PCR-based, genetic linkage map for human chromosome 16, and ultimately for the

**TABLE 2**

**Informativeness of Human Chromosome 16 $(AC)_n$ Repeat Loci**

| Locus | Repeat | Alleles[b] | Frequencies | Expected heterozygosity | PIC value |
|---|---|---|---|---|---|
| 16AC2.5 | | 170 | 0.01 | 0.79 (240)[c] | 0.76 |
| (D16S291) | P(25)[a] | 168 | 0.01 | | |
| | | 166 | 0.08 | | |
| | | 164 | 0.23 | | |
| | | 162 | 0.25 | | |
| | | 160 | 0.12 | | |
| | | 158 | 0.03 | | |
| | | 156 | 0.26 | | |
| | | 154 | 0.01 | | |
| 16AC2.3 | CP(18) | 198 | 0.01 | 0.74 (158) | 0.71 |
| (D16S292) | | 196 | 0.01 | | |
| | | 194 | 0.02 | | |
| | | 192 | 0.09 | | |
| | | 190 | 0.13 | | |
| | | 188 | 0.04 | | |
| | | 186 | 0.01 | | |
| | | 184 | 0.08 | | |
| | | 182 | 0.16 | | |
| | | 180 | 0.45 | | |
| 16AC1 | I(32) | 140 | 0.56 | 0.49 (156) | 0.37 |
| (D16S294) | max. P(7) | 138 | 0.44 | | |
| 16AC3.12 | I(33) | 192 | 0.01 | 0.79 (156) | 0.77 |
| (D16S298) | max. P(24) | 190 | 0.01 | | |
| | | 188 | 0.06 | | |
| | | 186 | 0.22 | | |
| | | 184 | 0.10 | | |
| | | 182 | 0.11 | | |
| | | 180 | 0.34 | | |
| | | 178 | 0.12 | | |
| | | 176 | 0.01 | | |
| | | 174 | 0.01 | | |
| | | 172 | 0.01 | | |
| 16AC6.17 | P(19) | 124 | 0.07 | 0.72 (158) | 0.69 |
| (D16S299) | | 122 | 0.03 | | |
| | | 120 | 0.20 | | |
| | | 118 | 0.16 | | |
| | | 116 | 0.45 | | |
| | | 114 | 0.07 | | |
| | | 112 | 0.01 | | |
| | | 110 | 0.01 | | |
| 16AC1.1 | CI(30) | 175 | 0.02 | 0.61 (156) | 0.57 |
| (D16S300) | max. P(24) | 173 | 0.04 | | |
| | | 171 | 0.15 | | |
| | | 169 | 0.08 | | |
| | | 167 | 0.60 | | |
| | | 165 | 0.07 | | |
| | | 163 | 0.04 | | |
| 16AC6.5 | I(25) | 111 | 0.99 | 0.02 (80) | 0.02 |
| (D16S302) | max. P(10) | 109 | 0.01 | | |
| 16AC1.14 | P(20) | 158 | 0.01 | 0.60 (152) | 0.58 |
| (D16S304) | | 152 | 0.02 | | |
| | | 150 | 0.01 | | |
| | | 148 | 0.02 | | |
| | | 146 | 0.06 | | |
| | | 144 | 0.03 | | |
| | | 142 | 0.11 | | |
| | | 140 | 0.61 | | |
| | | 138 | 0.08 | | |
| | | 136 | 0.02 | | |
| | | 134 | 0.02 | | |
| | | 130 | 0.01 | | |

entire human genome. In this regard it is worthwhile to compare the results presented here with those of Iizuka *et al.* (1992), where chromosome 11-specific *Alu*-SSCP polymorphisms were characterized. Of the 16 *Alu*-SSCP markers characterized, only 3 had PIC values >0.50 (only one > 0.70), and 7 markers were not polymorphic. While these *Alu*-SSCP polymorphisms will no doubt be a valuable new source of genetic markers, the AC repeat microsatellites, despite some difficulties in their characterization, are more likely to fulfill the role of PCR-based index markers for the human genome.

## ACKNOWLEDGMENTS

This work was supported by grants from the U.S. Department of Energy (Contract DEFG02 89ER60863) and the Australian National Health and Medical Research Institute. R.I.R. thanks Shelley Richards for support and encouragement during the preparation of the manuscript.

## REFERENCES

Bullock, W. O., Fernandez, J. M., and Short, J. M. (1987). XL-1 Blue: A high efficiency plasmid transforming rec A. *Escherchia coli* strain with β-galactosidase selection. *Biotechniques* **5:** 376.

Callen, D. F. (1986). A mouse–human hybrid cell panel for mapping human chromosome 16. *Ann. Genet.* **29:** 235–239.

Callen, D. F., Baker, E., Eyre, H. J., and Lane, S. A. (1990). An expanded mouse–human hybrid cell panel for mapping human chromosome 16. *Ann. Genet.* **33:** 190–195.

Cornall, R. J., Aitman, T. J., Hearne, C. M., and Todd, J. A. (1991). The generation of a library of PCR analyzed microsatellite variants for genetic mapping of the mouse genome. *Genomics* **10:** 874–881.

Hamada, H., Petrino, M. G., and Kakunaga, T. (1982). A novel repeated element with Z-DNA forming potential is widely found in evolutionarily diverse eukaryotic genomes. *Proc. Natl. Acad. Sci. USA* **79:** 6465–6469.

Iizuka, M., Mashiyama, S., Oshimura, M., Sekiya, T., and Hayashi, K. (1992). Cloning and polymerase chain reaction–single-strand conformation polymorphism analysis of anonymous *Alu* repeats on chromosome 11. *Genomics* **12:** 139–146.

Kogan, S. C., Doherty, M., and Gitschier, J. (1988). An improved method for prenatal diagnosis of genetic diseases by analysis of amplified DNA sequences: Application to hemophilia A. *New. Engl. J. Med.* **317:** 985–990.

Litt, M., and Luty, J. A. (1989). A hypervariable microsatellite revealed by *in vitro* amplification of a dinucleotide repeat within the cardiac muscle actin gene. *Am. J. Hum. Genet.* **44:** 397–401.

Luty, J. A., Guo, Z., Willard, H. F., Ledbetter, D. H., and Litt, M. (1990). Five polymorphic microsatellite VNTRs on the human X chromosome. *Am. J. Hum. Genet.* **46:** 776–783.

Reeders, S. T., and Hildebrand, C. E. (1989). Report of the committee on the genetic constitution of chromosome 16, Human Gene Mapping 10 (1989): Tenth International workshop on human gene mapping. *Cytogenet. Cell Genet.* **51:** 299–318.

Reeders, S. T., Hildebrand, C. E., and Sutherland, G. R. (1990). Report of the committee on the genetic constitution of chromosome 16, Human Gene Mapping 10.5 (1990): Update of the tenth international workshop on human gene mapping. *Cytogenet. Cell Genet.* **55:** 194–197.

Richards, R. I., Shen, Y., Holman, K., Kozman, H., Hyland, V. J., Mulley, J. C., and Sutherland, G. R. (1991a). Fragile X syndrome: Diagnosis using highly polymorphic microsatellite markers. *Am. J. Hum. Genet.* **48:** 1051–1057.

Richards, R. I., Holman, K., Shen, Y., Kozman, H., Harley, H., Brook, D., and Shaw, D. (1991b). Human glandular kallikrein genes: Genetic and physical mapping of the KLK1 locus using a highly polymorphic microsatellite PCR marker. *Genomics* **11:** 77–82.

Richards, R. I., Holman, K., Kozman, H., Kremer, E., Lynch, M., Pritchard, M., Yu, S., Mulley, J. C., and Sutherland, G. R. (1991c). Fragile X syndrome: Genetic localisation by linkage mapping of two microsatellite repeats of FRAXAC1 and FRAXAC2 which immediately flank the fragile site. *J. Med. Genet.* **28:** 818–823.

Richards, R. I., Holman, K. H., Lane, S., Sutherland, G. R., and Callen, D. F. (1991d). Human chromosome 16 physical map: Mapping of somatic cell hybrids using multiplex PCR deletion analysis of sequence tagged sites. *Genomics* **10:** 1047–1052.

Stallings, R. L., Ford, A. F., Nelson, D., Torney, D. C., Hildebrand, C. E., and Moyzis, R. K. (1991). Evolution and distribution of $(GT)_n$ repetitive sequences in mammalian genomes. *Genomics* **10:** 807–815.

Stallings, R. L., Torney, D. C., Hildebrand, C. E., Longmire, J. L., Deaven, L. L., Jett, J. H., Doggett, N. A., and Moyzis, R. K. (1990). Physical mapping of human chromosomes by repetitive sequence finger printing. *Proc. Natl. Acad. Sci. USA* **87:** 6218–6222.

Weber, J. L. (1990). Informativeness of human $(dC\text{-}dA)_n$ $(dG\text{-}dT)_n$ Polymorphisms. *Genomics* **7:** 524–530.

Weber, J. L., and May, P. E. (1989). Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am. J. Hum. Genet.* **44:** 388–396.

Shen, Y., Baker, E., Callen, D.F., Sutherland, G.R., Willson, T.A., Rakar, S. and Gough, N.M. (1992). Localization of the human GM-CSF receptor β chain gene (CSF2RB) to chromosome 22q12.2→q13.1. *Cytogenetics and Cell Genetics, 61(3), 175–177*.

Shen, Y., Thompson, A.T., Holman, K., Callen, D.F., Sutherland, G.R. and Richards, R.I. (1992). Four dinucleotide repeat polymorphisms on human chromosome 16 at D16S289, D16S318, D16S319 and D16S320. *Human Molecular Genetics*, 1(9), 773.

Schuffenhauer, S., Callen, D.F., Seidel, H., Shen, Y., Lederer, G. and Murken, J. (1992). De novo interstitial deletion 16 (q12. 1q13) of paternal origin in a 10-year-old boy. *Clinical Genetics*, 42(5), 246–250.

GENOMICS 13, 1178–1185 (1992)

# High-Resolution Cytogenetic-Based Physical Map of Human Chromosome 16

D. F. CALLEN,* N. A. DOGGETT,† R. L. STALLINGS,† L. Z. CHEN,* S. A. WHITMORE,* S. A. LANE,* J. K. NANCARROW,* S. APOSTOLOU,* A. D. THOMPSON,* N. M. LAPSYS,* H. J. EYRE,* E. G. BAKER,* Y. SHEN,* K. HOLMAN,* H. PHILLIPS,* R. I. RICHARDS,* AND G. R. SUTHERLAND*

*Department of Cytogenetics and Molecular Genetics, Adelaide Children's Hospital, King William Road, North Adelaide, South Australia, Australia 5006; and †Center for Human Genome Studies and Life Sciences Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545

Received May 7, 1992; revised May 11, 1992

**A panel of 54 mouse/human somatic cell hybrids, each possessing various portions of chromosome 16, was constructed; 46 were constructed from naturally occurring rearrangements of this chromosome, which were ascertained in clinical cytogenetics laboratories, and a further 8 from rearrangements spontaneously arising during tissue culture. By mapping 235 DNA markers to this panel of hybrids, and in relation to four fragile sites and the centromere, a cytogenetic-based physical map of chromosome 16 with an average resolution of 1.6 Mb was generated. Included are 66 DNA markers that have been typed in the CEPH pedigrees, and these will allow the construction of a detailed correlation of the cytogenetic-based physical map and the genetic map of this chromosome. Cosmids from chromosome 16 that have been assembled into contigs by use of repetitive sequence fingerprinting have been mapped to the hybrid panel. Approximately 11% of the euchromatin is now both represented in such contigs and located on the cytogenetic-based physical map. This high-resolution cytogenetic-based physical map of chromosome 16 will provide the basis for the cloning of genetically mapped disease genes, genes disrupted in cytogenetic rearrangements that have produced abnormal phenotypes, and cancer breakpoints.** 

## INTRODUCTION

The Human Genome Initiative will map and sequence the human genome. The maps of human chromosomes currently being constructed are either based on genetic distances estimated from meiotic recombination between polymorphic loci or based on the physical relationships between cloned DNA fragments. Various strategies have been utilized in determining physical relationships. These include the mapping of infrequent restriction sites by analysis of large DNA fragments generated by pulsed-field gel electrophoresis (Smith and Cantor, 1987) and the construction of contigs of overlapping DNA fragments cloned into cosmids or yeast artificial chromosomes (YACs) (Burke *et al.*, 1987). Radiation hybrid mapping, a statistical method based on the X-ray breakage of chromosomes, allows determination of order and distance between DNA markers in the megabase range (Cox *et al.*, 1990). These procedures allow the ordering of cloned fragments and the estimation of physical distances between them, but do not determine the location of cloned fragments on the chromosome. The construction of such cytogenetic-based maps can be achieved by *in situ* hybridization and by the use of rodent/human somatic cell hybrids that retain various segments of human chromosomes. Cytogenetic-based physical maps enable integration of the mapping information obtained using all other approaches. This integration is essential when considering approaches to cloning genetically mapped disease genes, genes disrupted by cytogenetic rearrangements, and cancer breakpoints.

Cytogenetic maps can be based entirely on *in situ* hybridization of cloned DNA fragments to metaphase chromosomes. Use of extended prophase chromosomes can provide a resolution of about 3 Mb, and *in situ* hybridization to interphase nuclei has been reported to allow ordering of cloned segments with a resolution of several hundred kilobases (Trask, 1991). These techniques cannot be readily used to map sequence-tagged sites (STSs) (Olson *et al.*, 1989), which are usually short stretches of DNA too small for use as probes for *in situ* hybridization. The resolution of *in situ* hybridization for chromosome mapping can be improved by utilizing metaphase chromosomes with defined breakpoints, for example, translocation breakpoints or fragile sites.

Somatic cell hybrid panels can also be used to generate cytogenetic-based physical maps. Such hybrids are generally constructed from a mouse or hamster cell line that is deficient for a gene selectable in tissue culture. When such a rodent line is fused with human cells and appropriate selection is applied, the rodent/human hy-



0888-7543/92 $5.00

brid rapidly eliminates human chromosomes until only one, or a subset, of the human chromosomes remains, including the human chromosome with the selectable gene. If the human parent line contains a chromosome rearrangement, panels containing various portions of a particular chromosome can be constructed. Such panels provide a method for the rapid localization of probes to chromosomal regions and are particularly suited to mapping STSs by PCR. The potential resolution of such panels is unlimited if it is assumed that each constitutional cytogenetic breakpoint ascertained in the human population has a unique location. Hybrid panels have been utilized for detailed mapping of specific chromosome regions, for example, in the vicinity of the fragile X at Xq27.3 (Suthers *et al.*, 1991) and in the vicinity of the WAGR syndrome complex at 11p13 (Couillin *et al.*, 1989). However, the resolution in the mapping of entire chromosomes has thus far been limited. Chromosome 3 has been divided into 11 intervals to give an average resolution of 17 Mb (Drabkin *et al.*, 1990), chromosome 13 into 6 intervals for a resolution of 18 Mb (Cowell and Mitchell, 1989), chromosome 17 into 13 intervals for a resolution of 6 Mb (van Tuinen *et al.*, 1987), and chromosome 21 into 11 intervals for a resolution of 5 Mb (Gao *et al.*, 1991). By generating an extensive somatic cell hybrid panel for chromosome 16, we have been able to construct a detailed cytogenetic-based physical map of this chromosome at an average resolution of less than 2 Mb.

## MATERIALS AND METHODS

*Mouse/human hybrids.* Mouse/human hybrids were derived by selecting for the gene *APRT* at 16q24. Mouse cell line A9 or 585MEL was fused with human cells using polyethylene glycol and hybrids selected by growth in medium containing alanosine and adenine as described (Callen, 1986). Human chromosomes in hybrids were identified cytogenetically by distamycin A/DAPI and G11 banding. Further characterization was by *in situ* hybridization to metaphase chromosomes with biotinylated total human DNA (to detect all human chromosomes) or chromosome 16 DNA (to detect chromosome 16-derived segments). Hybrids generally contained the portion of chromosome 16 from the breakpoint to the tip of the long arm in the absence of the normal 16 or any other chromosome 16 segments. The human parent was from fibroblasts, lymphoblastoid cell lines, or lymphocytes.

The majority of hybrids were derived from naturally occurring rearrangements of chromosome 16 ascertained in cytogenetics laboratories (Callen, 1986; Callen *et al.*, 1990b; Table 1). Exceptions were CY13A, CY18A, and CY180A, which arose spontaneously from the hybrids CY13, CY18, and CY180, respectively. Hybrids CY105, CY112, CY145, and CY199 had rearrangements of chromosome 16 that differed from the human parental line and occurred spontaneously during the fusion process. CY145 possessed two regions of chromosome 16, from p11.2 to p12.3 and from q23.1 to qter. Many of these spontaneously arising rearrangements involved insertion or translocation of chromosome 16 with mouse chromosomes. The chromosome 16 translocation in the human parent of CY198 spontaneously arose during tissue culture of the human fibroblast line. From each hybrid, DNA was extracted and analyzed by a multiplexed panel of STSs (Richards *et al.*, 1991) to confirm the integrity of the retained portion of chromosome 16.

*DNA probes.* The origin and descriptions of gene probes and anonymous DNA sequences that have been assigned "D16S" numbers are described in Reeders *et al.*, (1991) or are accessible from the Genome Data Base (Baltimore). Fingerprinted cosmids are from a library constructed from sorted chromosome 16 (Stallings *et al.*, 1990), and STSs have been constructed from a number of these cosmids (Doggett and Stallings, unpublished). Other anonymous probes were from a library derived from the hybrid CY3 (Hyland *et al.*, 1989). Oligo primers flanking $(AC)_n$ microsatellite repeats were as described (Weber *et al.*, 1990; Phillips *et al.*, 1991a,b; Shen *et al.*, 1991; Thompson *et al.*, 1992). In some cases, genes or anonymous DNA probes were converted to STS format by synthesis of appropriate oligoprimers (Table 2).

*Physical mapping to the hybrid panel.* DNA from the hybrid panel was analyzed by Southern blot in the case of gene probes, polymorphic probes that had been genetically mapped, and anonymous DNA probes derived from various chromosome 16-specific libraries or other sources. A number of STSs, including both polymorphic $(AC)_n$ microsatellite repeats and nonpolymorphic sequences, were regionally mapped by PCR analysis of the hybrid panel DNA.

*In situ* hybridization was used to map probes to either side of four fragile sites on chromosome 16. These were *FRA16A*, a rare folate-sensitive fragile site at 16p13.11; *FRA16B*, a rare distamycin A-inducible fragile site at the interface of band 16q21 and 16q22.1; and *FRA16D*, a common fragile site at 16q23.1 induced by aphidicoline treatment (Sutherland and Hecht, 1985). *FRA16E* is a distamycin A-inducible fragile site at 16p12.1, which has been found only in the Japanese population (Hori *et al.*, 1988). In addition, *in situ* hybridization of probes to high-resolution banded chromosomes was used to aid the aligning of breakpoints to the idiogram of chromosome 16.

*Physical mapping of cosmid contigs.* Cosmids from a chromosome 16-specific library have been assembled into contigs by use of repetitive sequence fingerprinting (Stallings *et al.*, 1990). Some of these contigs were physically mapped to the hybrid cell panel either by Southern blot analysis or by PCR (when sequence data were available). Alternatively, filters containing all of the approximately 4200 fingerprinted cosmids in high-density arrays were screened by hybridization with probes of known physical location (Stallings *et al.*, 1992b). This allowed the identification, and therefore determination of the physical location, of the cosmid contig containing the region complementary to the probe.

## RESULTS

Because of the number of breakpoints represented in the mouse/human hybrid panel, it was not always possible to order them by the cytogenetic analysis of the parental line. Unequivocable cytogenetic definition of breakpoints is often difficult due to the juxtaposition of similar banding regions of the chromosome or the poor resolution of banding. In many instances, cytogenetic breakpoints were found to be located in regions that differed from their original description. For example, cytogenetic analysis had assigned many of the translocation breakpoints represented in hybrids located distal to *FRA16D* (and therefore at or distal to 16q23.1) to 16q22.

A total of 235 DNA markers on chromosome 16 were mapped against this cell panel of 54 hybrids containing various portions of chromosome 16. These markers include 26 genes, 17 of which are in an STS format; 71 fingerprinted cosmids, 38 of which are STSs; and 106 other anonymous DNA probes, 16 of which are STSs. There were 40 $(AC)_n$ microsatellite repeats, 3 isolated from previously mapped fingerprinted cosmids and 5 from previously mapped DNA markers. Included in these 235 DNA markers are 66 polymorphic markers,

**TABLE 1**

**Origin of Mouse/Human Hybrids**

| Human parent line | | | | Hybrid | |
|---|---|---|---|---|---|
| Name | Chromosome 16 rearrangement[a] | Cell type | Source | Name | 16 material present |
| GM8352 | t(6;16)(q15;q24) | Lymphoblast | NIGMS[b] | CY100 | der(6)t(6;16) |
| S2787 | t(16;19)(q23;q13.4) | Amniocyte | M. Mikkelson, Denmark | CY110 | der(19)t(16;19) |
| GM9132 | t(9;16)(p22;q23.1) | Fibroblast | NIGMS | CY115 | der(9)t(9;16) |
| TR2 | t(5;16)(q14.2;q22.1) | Lymphoblast | D.F. Callen, Australia (Callen *et al.*, 1988) | CY116 | der(5)t(5;16) |
| TR3 | t(2;16)(p15;q22.3) | Lymphoblast | D.F. Callen, Australia (Callen *et al.*, 1988) | CY117 | der(2)t(2;16) |
| G7612/90 | t(5;16)(q22.1;q22.1) | Aminocyte | S. Purvis-Smith, Australia | CY121 | der(5)t(5;16) |
| SOM | t(12;16)(q13;q22) | Amniocyte | K. Dyer, Vivigen Inc., U.S.A. | CY122 | der(12)t(12;16) |
| DD | t(15;16)(q26.1;q22.1) | Fibroblast | D. Marsden, U.S.A. (Nyhan *et al.*, 1989) | CY124 | der(15)t(15;16) |
| YEM | del(16)(q22.1) | Fibroblast | D.F. Callen, Australia | CY127 | del(16)(q22.1) |
| FP945 | del(16)(q12) | Fibroblast | J. Murken, S. Schuffenhauer, Germany | CY138 | del(16)(q12) |
| 770815 | t(2;16)(q11;q13 or q12) | Amniocyte | A. Fujimoto, U.S.A. | CY148 | der(2)t(2;16) |
| ZYL | t(1;16)(p11;p11.1) | Amniocyte | G. Peters, Australia | CY152 | der(16)t(1;16) |
| GM6316A | t(11;16)(p11.2;p11.1) | Lymphoblast | NIGMS | CY153 | der(16)t(11;16) |
| 430/85RII | t(5;15;16)(q35.3;q11.2;p11.1) | Fibroblast | M. Fraccaro, Italy | CY155 | der(16)t(5;15;16) |
| NEZ | t(6;16)(q27;p12.1) | Fibroblast | R. Haluza, Czechoslovakia | CY163 | der(16)t(6;16) |
| TC126 | ins(11;16)(q13;p13.13p13.13) | Fibroblast | J.A. Bell, Australia (Callen *et al.*, 1990a) | CY180 | der(16)ins(11;16) |
| 31.2 | t(7;16)(p22;p13) | Lymphoblast | M.H. Breuning, The Netherlands | CY183 | der(16)t(7;16) |
| 1019-89BL | t(4;16)(q25;p13) | Lymphoblast | M. Mikkelson, Denmark | CY190 | der(16)t(4;16) |
| ROM | t(1;16)(q23;p13) | Fibroblast | G. Peters, Australia | CY198 | der(16)t(1;16) |

[a] The breakpoints are as given in their initial description.
[b] NIGMS: Human Genetic Mutant Cell Repository, Camden, NJ.

including both RFLPs and $(AC)_n$ microsatellite repeats, which have been typed in the CEPH pedigrees.

The complete set of results can be represented as a two-way array, with the chromosome 16 breakpoints present in somatic cell hybrids and fragile sites on one axis and DNA markers on the other axis. Along each axis the breakpoints or markers can be unambiguously ordered from the tip of the short arm to the tip of the long arm of chromosome 16. Four fragile sites, the centromere, and the localization of four probes (*D16S60, D16S333, D16S47,* and *D16S344*) by high-resolution *in situ* hybridization were used to link the order of breakpoints and probes to the idiogram of chromosome 16 (Fig. 1). The cytogenetic location of fragile sites can be accurately determined as there is a break on a chromosome 16 that can be precisely defined using high-resolution banding.

A total of 52 intervals on chromosome 16 could be defined by these data. There were nine instances in which two, and two instances in which three, different breakpoints could not be resolved since no DNA markers have been mapped between them. Each fingerprinted cosmid mapped is a member of a contig generated by repetitive sequence fingerprinting. In addition, cosmid contigs corresponding to a further 21 physically mapped probes were identified by hybridization to filters containing the cosmids in high-density arrays. By this procedure, 9.3 Mb of DNA cloned into cosmid contigs has been physically mapped.

The results from the mapping of two clones were complex. The locus *D16S79* mapped to both sides of *FRA16A* by *in situ* hybridization. The 1.0-kb probe 36.1 detects two *Taq*I polymorphisms at this locus in genomic DNA, locus A with alleles at 1.6 and 3.4 kb and locus B with alleles at 4.5 and 9.5 kb. There were no *Taq*I sites present in probe 36.1. A cosmid clone, c66F3, was isolated by probing high-density arrays of cosmids with 36.1. When this cosmid c66F3 was digested with *Taq*I, there was a single 4.5-kb band detected with 36.1. These results suggest that genomic DNA contains two copies of the 36.1 sequence that are on either side of *FRA16A*, with locus *D16S79A* detecting one polymorphism and *D16S79B* detecting the other. Genetic linkage studies (Kozman *et al.*, unpublished) are consistent with *D16S79A* being distal to *D16S79B*. The size of the duplicated segments detected by 36.1 and their distance apart

**TABLE 2**

**STSs Physically Mapped on Chromosome 16**

| Locus | Primers[a] | PCR product (bp) | Reference[b] |
|---|---|---|---|
| *ALDOA* | CTC CAG AGA ATC AGA ACA GCC ACC A<br>CTG GTT ATA AGG CAG CTC CTG CCC G | 475 | Izzo *et al.* (1987) |
| *CA7* | CTC TCG GCT GCA GAT GGA TCC TCT G<br>CAG GCC CAG AAC ATT CAT GTG TAC A | 237 | Montgomery *et al.* (1991) |
| *DPEP1* | AGA CCA GAG TCC CCT TTA GGG TTC C<br>GTG TTG CTT TTA TTG CGG GCC TAC T | 194 | Adachi *et al.* (1990) |
| *GNAO1* | AAT CGA TTG CTC CGA CCA ATC CAC T<br>CCC TAG AAC ATT TTG ACT ACT GCA GG | 359 | Murtagh *et al.* (1991) |
| *GOT* | GTC TTC AGC CTC TGC TAT TGA GAG C<br>AAC CAG GCC GCA ATA AGT CTA CAT G | 1000 | Pol *et al.* (1988) |
| *MPG* | ATG CCC GTG CAG CTC GCA CAT ATG<br>CTG CCC TGC TCT AGC TGG TCG CTG | 300 | Samson *et al.* (1991) |
| *D16S79*(36.1) | TAC TGT GAA TTA GGT TAA ACA TTG TTT GAA GG<br>TCA AAT AGT GTT TAC TGA ATC TTT GAT CTG CT | 926 | |
| *D16S94*(VK5) | GAG CCC ACA GAC TTC TTT CCT GGG<br>CCC ACA CCA TGT ACA GCC AGT AGG TA | 86 | |
| *D16S96*(VK20) | GGT TCT TGG CCA AGA TGA GGC TGA A<br>AAC CAG GAA CAC TGC CTT TCC TAC | 180 | |
| *D16S163*(16-10) | GTC CTG CTC ACC AGC AAG TGC TGA A<br>AAC CTA GCT AAT TCC ATC AAG GAG G | 333 | |
| *D16S174*(16-53) | GGC CTT CTA GTG CCA TTT CTA AAT G<br>CCA GAC TCT TTC TGC GTA ACT TTC C | 296 | |
| *D16S265*(MFD23) | TCT GAG TCC TCT GTG CAC TTT GTG<br>GGA ATT TGG GAA GCA TGG TGT CTC | 160 | |
| *D16S317*(16AC7.9) | AGG AGT ACA CAG ATG CAC CAG ACC T<br>CCT TTG ATA CCC TGG CTT TGC AGT C | 136–138 | |
| *D16S321*(T118) | AGG CCT CCA AAG ATG TTT<br>TTA GCT GAG CGT GGT GGT | 201 | |
| *D16S322*(T117) | AAG ACC AGC CTG GCC AAC AT<br>GAG ACA GTG TTT CTA TCT CC | 196 | |
| *D16S323*(T116) | CTG AAA GAC ATC GAA ACA CTC C<br>GCA GCT GGG GAA CAG ACT TTA T | 154 | |
| *D16S325*(T301) | AAC ATG ATG AAA CCC CGT CT<br>TCA GCT CAC TGC AAC CTC TG | 220 | |
| *D16S326*(T210) | TTC CTA GTG CAT AGT CTG GCT GCC T<br>GCC TAG GCA ACA GAG TGA GA | 120 | |
| *D16S397*(MFD98) | GCA ACT ACC TCC AAG TCT GGA GAC CAC TAA<br>GAA TTG GAT TAT AGG CAT GAG CCA CTG CAC | 187 | |
| *D16S398*(MFD168) | CTT CCA ACT CAT TCC CTG TGT GAC TTT AGG<br>CCA GAA GAA ACC AAG TGG GTT AGG TCA GGA | 230 | |
| (16AC7.22) | ACC ATG CCC AGC GTA GGT AAT T<br>ATC TAC AGA AAT GTT TGA GAC C | 168 | |
| (16AC7.59) | CTG AAT ATT TCT GTG CCT AGG TTT C<br>CAA CAG AGT GAG ACC CTG TCT CTA TA | 175 | |

[a] Primers are given in the orientation 5′ to 3′ with the forward primer given first. The primer sequences for STSs from 11 additional genes are given in Richards *et al.* (1991); for *D16S164*(AC16-15), *D16S168*(AC16-42), and *D16S186*(AC16-101) in Phillips *et al.* (1991a); for *D16S287*(AC16XE81) in Phillips *et al.* (1991b); and for *D16S290*(T102) in Chen *et al.* (1991). Sequences for STSs derived from fingerprinted cosmids are given in Doggett *et al.* (in preparation). Further primers that flank $(AC)_n$ repeats have been published in Thompson *et al.* (1992) and in Shen *et al.* (1991, and unpublished).

[b] References are given when primer sequences were derived from a published gene sequence. Primer sequences for *D16S260*(MFD12), *D16S261*(MFD24), *D16S266*(MFD62), and *D16S267*(MFD65) are as described (Webber *et al.*, 1990). The primers used for *D16S265*(MFD23), *D16S397*(MFD98), and *D16S398*(MFD168) are as given.

are unknown, although analysis of rare restriction sites suggests that this distance is greater than 1 Mb.

In addition to this duplication, cosmids c13E7, c307G6, and c306D6 were found to map to regions on both the p and q arms of this chromosome. These clones come from the same cosmid contig (55) known to contain a new type of minisatellite repetitive sequence (Stallings *et al.*, 1992b).

The results of mapping the locus *D16S57* to the hybrid cell panel were not consistent. For example, the hybrids CY123, CY15, CY165, CY155, CY160, and CY153 showed hybridization to this probe, which CY13, CY180A, CY152, and CY199 did not. Fluorescence *in situ* hybridization established that in some individuals this locus was missing on one of the chromosome 16 homologs, and a complementary sequence was consis-

homologs, and a complementary sequence was consistently located on the tip of the short arm of chromosome 6.

## DISCUSSION

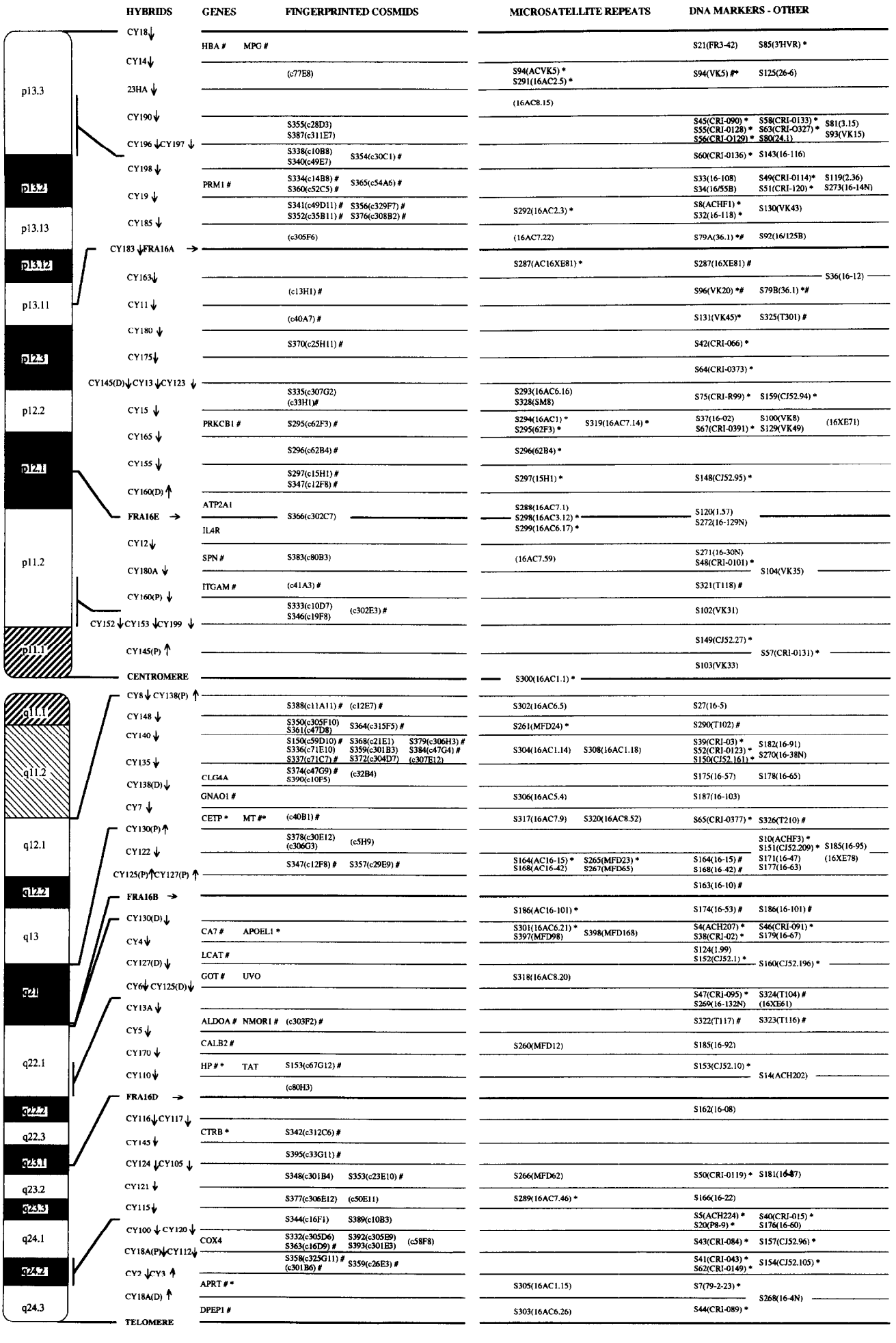
The mapping of 235 DNA markers to a panel of 54 somatic cell hybrids, four fragile sites, and the centromere allowed the delineation of 52 intervals on chromosome 16. Assuming that all the breakpoints of chromosome 16 represented in this panel are unique, there are potentially 67 intervals. Clones containing the pericentric alphoid or satellite repeats have not been mapped to this panel. No unique clones have been mapped in the region from the centromere to the cluster of breakpoints defined by CY8 and CY138(P), a region that encompasses these repetitive sequences. The euchromatic portion of chromosome 16, 85 Mb (Stallings *et al.*, 1990), has therefore been divided into 52 regions, which corresponds to an average resolution of the physical map of 1.6 Mb. An independent estimation of this resolution is provided by the interstitial deletion of 16q22.1, which is represented in CY125. This deletion has been estimated to be 7 Mb by flow karyotypic analysis (Cooke *et al.*, 1987). Four intervals have been defined between the breakpoints of this deletion by hybrid breakpoints and *FRA16A*, an average resolution of 1.75 Mb. Mapping of cosmids that are members of contigs assembled by repetitive sequence fingerprinting and identification of such contigs corresponding to mapped probes (Stallings *et al.*, 1992a) have allowed the mapping of 9.3 Mb of cloned DNA. Therefore, approximately 11% of the euchromatin of chromosome 16 has now been cloned into cosmids, and these cosmids located to a detailed cytogenetic-based physical map.

The short arm of chromosome 16 has regions that are duplicated. This study shows that the region encompassing *D16S79* is duplicated on either side of the fragile site, *FRA16A*. The locus *D16S125*, at 16p13.3 in the vicinity of the polycystic kidney disease (*PKD1*) gene, has also been shown to be duplicated (Reeders *et al.*, 1990). In addition, there are regions of the short arm of chromosome 16, which are missing from some individuals. The cloned minisatellite, MS29 (Wong *et al.*, 1990), shows a similar localization on chromosomes 6 and 16 as does *D16S57* with a duplicated segment on chromosome 16 being absent in some individuals. It is not known whether *D16S57* and MS29 are located on the same duplicated segment. There is also a telomere length polymorphism involving differences between chromosome 16s of up to 260 kb of DNA (Wilkie *et al.*, 1991). The comparative genetic map between mouse and human (Davisson *et al.*, 1991) suggests that while the long arm of chromosome 16 is syntenic with a single mouse chromosome (MMU8), the short arm of human 16 has synteny with four different mouse chromosomes (MMU7, 11, 16, and 17). It is of interest that the tip of the short arm of human chromosome 6 and the proximal portion of the short arm of 16 both show homology with the same mouse chromosome (MMU7), and this is the region that has common sequences (MS29 and *D16S57*) in humans. These observations may reflect the evolutionary history of the short arm of chromosome 16, which must have involved several chromosome rearrangements.

Band 16q21 is the most intensely staining positive G-band of chromosome 16. There is an absence of localized genes and few hybrid breakpoints within this band (Fig. 1). A family that segregates for a deletion of this band with no phenotypic effect has been reported (Witt *et al.*, 1988), suggesting that it contains little DNA of genetic importance. The other positive G-bands on chromosome 16 are less intensely staining and do not show a comparable lack of breakpoints and gene localizations. These results are similar to those reported for chromosome 21 where the intensely staining G-positive-band 21q21 shows a paucity of breakpoints and genes and in addition lacks rare restriction enzyme-cutting sites (Gardiner *et al.*, 1990).

The high-resolution cytogenetic-based physical map of chromosome 16 will form the basis for the positional cloning of regions of interest, since any interval is on average less than 2 Mb, and for establishing correlations between the genetic and physical maps. Included in this physical map are 66 probes that have been typed in the CEPH families, allowing the construction of genetic maps of this chromosome (Julier *et al.*, 1990; Keith *et al.*,

**FIG. 1.** Cytogenetic-based physical map of human chromosome 16. The portion of chromosome 16 present in each mouse/human hybrid cell line is delineated by a horizontal line with an arrow indicating the direction of the retained portion of chromosome 16 from the breakpoint. Breakpoints of fragile sites are indicated by horizontal arrows. The majority of hybrid cell lines contain the portion of chromosome 16 from the breakpoint to the tip of the long arm (qter) and have been derived from translocations involving chromosome 16. Exceptions are five interstitial deletions (CY125, CY127, CY130, CY138, CY160); the hybrid CY18, which contains an intact chromosome 16; the hybrid CY180, which was derived from an insertional translocation of 16p13 into 11q; and CY18A and CY145, which contain one and two fragments of 16, respectively. In these instances where there are two breakpoints on the same arm (P) refers to the proximal, and (D) the distal breakpoint. Hybrids CY2 and CY3 contain the der(X) and the der(16), respectively, from the same t(X;16) translocation. DNA markers are divided into columns corresponding to genes, cosmids that have been assembled into contigs by repetitive sequence fingerprinting, oligonucleotide primers that flank $(AC)_n$ microsatellite repeats, and other anonymous DNA markers. When such $(AC)_n$ repeats have been identified in previously mapped DNA markers, the location of these markers is also given. All genes, cosmids or other DNA markers that have been converted to STS format by synthesis of oligonucleotide primers, are indicated by #. All polymorphic markers that have been typed in the CEPH pedigrees are indicated by *. The D numbers are abbreviated, e.g., *D16S3* to S3, and DNA marker names are given in brackets. *In situ* hybridization to prophase banded chromosomes provided additional links between the ordered panel of hybrid breakpoints and the idiogram. The location of the hybridization signal is indicated by a vertical line, and the DNA markers used are *D16S60* (16p13.3); *D16S333* (16p11.2); *D16S47* (16q22.1); and *D16S344* (16q24.2). The relative sizes of the idiograms for the short and long arms are not to scale. Note: The proximal breakpoint of CY127 has only been evaluated with PCR-based markers in the interval between CY130 (P) and CY125 (P).


HYBRIDS
GENES
FINGERPRINTED COSMIDS
MICROSATELLITE REPEATS
DNA MARKERS - OTHER
p13.3
p13.2
p13.13
p13.12
p13.11
p12.3
p12.2
p12.1
p11.2
p11.1
q11.1
q11.2
q12.1
q12.2
q13
q21
q22.1
q22.2
q22.3
q23.1
q23.2
q23.3
q24.1
q24.2
q24.3
CY18
CY14
23HA
CY190
CY196
CY197
CY198
CY19
CY185
CY183
FRA16A
CY163
CY11
CY180
CY175
CY145(D)
CY13
CY123
CY15
CY165
CY155
CY160(D)
FRA16E
CY12
CY180A
CY160(P)
CY152
CY153
CY199
CY145(P)
CENTROMERE
CY8
CY138(P)
CY148
CY140
CY135
CY138(D)
CY7
CY130(P)
CY122
CY125(P)
CY127(P)
FRA16B
CY130(D)
CY4
CY127(D)
CY6
CY125(D)
CY13A
CY5
CY170
CY110
FRA16D
CY116
CY117
CY145
CY124
CY105
CY121
CY115
CY100
CY120
CY18A(P)
CY112
CY2
CY3
CY18A(D)
TELOMERE
HBA #
MPG #
(c77E8)
S355(c28D3)
S387(c311E7)
S338(c10B8)
S340(c49E7)
S354(c30C1) #
PRM1 #
S334(c14B8) #
S360(c52C5) #
S365(c54A6) #
S341(c49D11) #
S352(c35B11) #
S356(c329F7) #
S376(c308B2) #
(c305F6)
(c13H1) #
(c40A7) #
S370(c25H11) #
S335(c307G2)
(c33H1)#
PRKCB1 #
S295(c62F3) #
S296(c62B4) #
S297(c15H1) #
S347(c12F8) #
ATP2A1
S366(c302C7)
IL4R
SPN #
S383(c80B3)
ITGAM #
(c41A3) #
S333(c10D7)
S346(c19F8)
(c302E3) #
S388(c11A11) #
(c12E7) #
S350(c305F10)
S361(c47D8)
S364(c315F5) #
S150(c59D10) #
S336(c71E10)
S337(c71C7) #
S368(c21E1)
S359(c301B3)
S372(c304D7)
S379(c306H3) #
S384(c47G4) #
(c307E12)
S374(c47G9) #
S390(c10F5)
(c32B4)
CLG4A
GNAO1 #
CETP *
MT #*
(c40B1) #
S378(c30E12)
(c306G3)
(c5H9)
S347(c12F8) #
S357(c29E9) #
CA7 #
APOEL1 *
LCAT #
GOT #
UVO
ALDOA #
NMOR1 #
(c303F2) #
CALB2 #
HP # *
TAT
S153(c67G12) #
(c80H3)
CTRB *
S342(c312C6) #
S395(c33G11) #
S348(c301B4)
S353(c23E10) #
S377(c306E12)
(c50E11)
S344(c16F1)
S389(c10B3)
COX4
S332(c305D6)
S363(c16D9) #
S392(c305E9)
S393(c301E3)
(c58F8)
S358(c325G11) #
(c301B6) #
S359(c26E3) #
APRT # *
DPEP1 #
S21(FR3-42)
S85(3'HVR) *
S94(ACVK5) *
S291(16AC2.5) *
S94(VK5) #*
S125(26-6)
(16AC8.15)
S45(CRI-090) *
S55(CRI-0128) *
S56(CRI-0129) *
S58(CRI-0133) *
S63(CRI-0327) *
S80(24.1)
S81(3.15)
S93(VK15)
S60(CRI-0136) *
S143(16-116)
S33(16-108)
S34(16/55B)
S49(CRI-0114)*
S51(CRI-120) *
S119(2.36)
S273(16-14N)
S292(16AC2.3) *
S8(ACHF1) *
S32(16-118) *
S130(VK43)
(16AC7.22)
S79A(36.1) *#
S92(16/125B)
S287(AC16XE81) *
S287(16XE81) #
S36(16-12)
S96(VK20) *#
S79B(36.1) *#
S131(VK45)*
S325(T301) #
S42(CRI-066) *
S64(CRI-0373) *
S293(16AC6.16)
S328(SM8)
S75(CRI-R99) *
S159(CJ52.94) *
S294(16AC1) *
S295(62F3) *
S319(16AC7.14) *
S37(16-02)
S67(CRI-0391) *
S100(VK8)
S129(VK49)
(16XE71)
S296(62B4) *
S297(15H1) *
S148(CJ52.95) *
S288(16AC7.1)
S298(16AC3.12) *
S299(16AC6.17) *
S120(1.57)
S272(16-129N)
(16AC7.59)
S271(16-30N)
S48(CRI-0101) *
S104(VK35)
S321(T118) #
S102(VK31)
S149(CJ52.27) *
S57(CRI-0131) *
S103(VK33)
S300(16AC1.1) *
S302(16AC6.5)
S27(16-5)
S261(MFD24) *
S290(T102) #
S304(16AC1.14)
S308(16AC1.18)
S39(CRI-03) *
S52(CRI-0123) *
S150(CJ52.161) *
S182(16-91)
S270(16-38N)
S175(16-57)
S178(16-65)
S306(16AC5.4)
S187(16-103)
S317(16AC7.9)
S320(16AC8.52)
S65(CRI-0377) *
S326(T210) #
S10(ACHF3) *
S151(CJ52.209) *
S171(16-47)
S177(16-63)
S185(16-95)
(16XE78)
S164(AC16-15) *
S168(AC16-42)
S265(MFD23) *
S267(MFD65)
S164(16-15) #
S168(16-42) #
S163(16-10) #
S186(AC16-101) *
S174(16-53) #
S186(16-101) #
S301(16AC6.21) *
S397(MFD98)
S398(MFD168)
S4(ACH207) *
S38(CRI-02) *
S46(CRI-091) *
S179(16-67)
S124(1.99)
S152(CJ52.1) *
S160(CJ52.196) *
S318(16AC8.20)
S47(CRI-095) *
S269(16-132N)
S324(T104) #
(16XE61)
S322(T117) #
S323(T116) #
S260(MFD12)
S185(16-92)
S153(CJ52.10) *
S14(ACH202)
S162(16-08)
S266(MFD62)
S50(CRI-0119) *
S181(16-87)
S289(16AC7.46) *
S166(16-22)
S5(ACH224) *
S20(P8-9) *
S40(CRI-015) *
S176(16-60)
S43(CRI-084) *
S157(CJ52.96) *
S41(CRI-043) *
S62(CRI-0149) *
S154(CJ52.105) *
S305(16AC1.15)
S7(79-2-23) *
S268(16-4N)
S303(16AC6.26)
S44(CRI-089) *

1990; Kozman *et al.*, 1992). A comprehensive correlation of the cytogenetic-based physical map and the genetic map of this chromosome can now be undertaken. This will be further refined by the use of highly informative $(AC)_n$ microsatellite repeats (Weber *et al.*, 1990; Phillips *et al.*, 1991a,b; Shen *et al.*, 1991). The mapped DNA markers in any one interval can be used to identify cosmid contigs (Stallings *et al.*, 1992a) or YACs enabling rapid isolation of cloned DNA.

In conclusion, the construction of a mouse/human hybrid cell panel of chromosome 16 has allowed division of this chromosome into 52 actual and 67 potential intervals and has allowed the physical mapping of 11% of the chromosome cloned into cosmid contigs. The mapping of cosmid contigs (Stallings *et al.*, 1990) and YACs will facilitate closure of contig gaps and enable a complete map of cloned DNA of chromosome 16 to be constructed. The correlation of the genetic map and the cytogenetic-based physical map will provide the opportunity to positionally clone disease genes on human chromosome 16. Evidence from nonrandom chromosome breaks in neoplasia (Mitelman *et al.*, 1991) and loss of heterozygosity in solid tumors (Tsuda *et al.*, 1990; Sato *et al.*, 1991; Carter *et al.*, 1990) suggest that there are specific regions of chromosome 16 involved in neoplasia. Use of this high-resolution cytogenetic-based physical map will facilitate cloning of such regions. There has been increased emphasis on the generation of sequences from randomly isolated cDNAs that can then be used as expressed sequenced tags (Adams *et al.*, 1991). Detailed physical mapping of ESTs will provide rapid access to candidate gene sequences in the vicinity of localized genes or chromosome breakpoints.

## ACKNOWLEDGMENTS

The cooperation of the many clinical cytogenetics laboratories that contributed cell lines possessing chromosome 16 rearrangements is greatly appreciated. In particular, we thank J. A. Bell, M. H. Breuning, K. Dyer, M. Fraccaro, A. Fujimoto, R. Haluza, T. A. Hori, D. Marsden, M. Mikkleson, J. Murken, G. Peters, S. Purvis-Smith, and S. Schuffenhauer for providing cell lines and Eric Kremer for STS sequence data. In addition we thank the laboratories that have generously provided gene and anonymous DNA markers. We acknowledge support from the United States Department of Energy Grant DE-FG02-89ER60863 and the Australian National Health and Medical Research Council.

## REFERENCES

Adachi, H., Tawaragi, Y., Inuzuka, C., Kubota, I., Tsujimoto, M., Nishihara, T., and Nakazato, H. (1990). Primary structure of human microsomal dipeptidase deduced from molecular cloning. *J. Biol. Chem.* **265:** 3992–3995.

Adams, M. D., Kelley, J. M., Gocayne, J. D., Dubnick, M., Polymeropoulos, M. H., Xiao, H., Merril, C. R., Wu, A., Olde, B., Moreno, R. F., Kerlavage, A. R., McCombie, W. R., and Venter, J. C. (1991). Complementary DNA sequencing: Expressed sequence tags and human genome project. *Science* **252:** 1651–1656.

Burke, D. T., Carle, G. F., and Olson, M. V. (1987). Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science* **236:** 806–812.

Callen, D. F. (1986). A mouse–human hybrid cell panel for mapping human chromosome 16. *Ann. Genet.* **29:** 235–239.

Callen, D. F., Baker, E., Eyre, H. J., Chernos, J. E., Bell, J. A., and Sutherland, G. R. (1990a). Reassessment of two apparent deletions of chromosome 16p to an ins(11;16) and a t(1;16) by chromosome painting. *Ann. Genet.* **33:** 219–221.

Callen, D. F., Baker, E., Eyre, H. J., and Lane, S. A. (1990b). An expanded mouse–human hybrid cell panel for mapping human chromosome 16. *Ann. Genet.* **33:** 190–195.

Callen, D. F., Hyland, V. J., Baker, E. G., Fratini, A., Simmers, R. N., Mulley, J. C., and Sutherland, G. R. (1988). Fine mapping of gene probes and anonymous DNA fragments to the long arm of chromosome 16. *Genomics* **2:** 144–153.

Carter, B. S., Ewing, C. M., Ward, W. S., Treiger, B. F., Aalders, T. W., Schalken, J. A., Epstein, J. I., and Isaacs, W. B. (1990). Allelic loss of chromosomes 16q and 10q in human prostate cancer. *Proc. Natl. Acad. Sci. USA* **87:** 8751–8755.

Chen, L. Z., Shen, Y., Holman, K., Thompson, A., Lane, S., Richards, R. I., Sutherland, G. R., and Callen, D. F. (1991). An STS at the *D16S290* locus. *Nucleic Acids Res.* **19:** 5793.

Cooke, A., Tolmie, J., Darlington, W., Boyd, E., Thomson, R., and Ferguson-Smith, M. A. (1987). Confirmation of a suspected 16q deletion in a dysmorphic child by flow karyotypic analysis. *J. Med. Genet.* **24:** 88–92.

Couillin, P., Azoulay, M., Henry, I., Ravise, N., Grisard, M. C., Jeanpierre, C., Barichard, F., Metezeau, P., Candelier, J. J., Lewis, W., Van Heyningen, V., and Jenien, C. (1989). Characterization of a panel of somatic cell hybrids for subregional mapping along 11p and within band 11p13. *Hum. Genet.* **82:** 171–178.

Cowell, J. K., and Mitchell, C. D. (1989). A somatic cell hybrid mapping panel for regional assignment of human chromosome 13 DNA sequences. *Cytogenet. Cell Genet.* **52:** 1–6.

Cox, D. R., Burmeister, M., Price, E. R., Kim, S., and Myers, R. M. (1990). Radiation hybrid mapping: A somatic cell genetic method for constructing high resolution maps of mammalian chromosomes. *Science* **250:** 245–250.

Davisson, M. T., Lalley, P. A., Peters, J., Doolittle, D. P., Hillyard, A. L., and Searle, A. G. (1991). Human Gene Mapping 11. Report of the comparative committee for human, mouse and other rodents. *Cytogenet. Cell Genet.* **58:** 1152–1189.

Drabkin, H., Wright, M., Jonsen, M., Varkony, T., Jones, C., Sage, M., Gold, S., Morse, H., Mendez, M., and Erickson, P. (1990). Development of a somatic cell hybrid mapping panel and molecular probes for chromosome 3. *Genomics* **8:** 435–446.

Gao, J., Erickson, P., Patterson, D., Jones, C., and Drabkin, H. (1991). Isolation and regional mapping of *Not*I and *Eag*I clones from human chromosome 21. *Genomics* **10:** 166–172.

Gardiner, K., Horisberger, M., Kraus, J., Tantravahi, U., Korenberg, J., Rao, V., Reddy, S., and Patterson, D. (1990). Analysis of human chromosome 21: Correlation of physical and cytogenetic maps; gene and CpG island distributions. *Embo. J.* **9:** 25–34.

Hori, T., Takahashi, E., Ishihara, T., Minamihisamatsu, M., Kaneko, Y., and Murata, M. (1988). Distamycin A-inducible fragile sites and cancer proneness. *Cancer Genet. Cytogenet.* **34:** 177–187.

Hyland, V. J., Fernandez, K. E. W., Callen, D. F., MacKinnon, R. N., Baker, E., Friend, K., and Sutherland, G. R. (1989). Assignment of anonymous DNA probes to specific intervals of human chromosomes 16 and X. *Hum. Genet.* **83:** 61–66.

Izzo, P., Costanzo, P., Lupo, A., Rippa, E., Borhese, A. M., Paolella, G., and Salvatore, F. (1987). A new human species of aldolase A mRNA from fibroblasts. *Eur. J. Biochem.* **164:** 9–13.

Julier, C., Nakamura, Y., Lathrop, M., O'Connell, P., Leppert, M., Mohandas, T., Lalouel, J-M., and White, R. (1990). A primary map of 24 loci on human chromosome 16. *Genomics* **6:** 419–427.

Keith, T. P., Green, P., Reeders, S. T., Brown, V. A., Phipps, P., Bricker, A., Falls, K., Rediker, K. S., Powers, J. A., Hogan, C., Nelson, C., Knowlton, R., and Donis-Keller, H. (1990). Genetic linkage

map of 46 DNA markers on human chromosome 16. *Proc. Natl. Acad. Sci. USA* **87:** 5754–5758.

Kozman, H. M., Phillips, H. A., Callen, D. F., Sutherland, G. R., and Mulley, J. C. (1992). Synthesis of the cytogenetic and genetic maps for human chromosome 16. Submitted for publication.

Mitelman, F., Kaneko, Y., and Trent, J. (1991). Human Gene Mapping 11. Report of the Committee on chromosome changes in neoplasia. *Cytogen. Cell Genet.* **58:** 1053–1079.

Montgomery, J. C., Venta, P. J., Eddy, R. L., Fukushima, Y-S., Shows, T. B., and Tashian, R. E. (1991). Characterization of the human gene for a newly discovered carbonic anhydrase, CAVII, and its localization to chromosome 16. *Genomics* **11:** 835–848.

Murtagh, J. J., Eddy, R., Shows, T. B., Moss, J., and Vaughan, M. (1991). Different forms of $G_{\alpha_o}$mRNA arise by alternative splicing of transcripts from a single gene on human chromosome 16. *Mol. Cell Biol.* **11:** 1146–1155.

Nyhan, W. L., Mascarello, J., Barshop, B., Doroski, D., and Hirschhorn, K. (1989). Duplication of 16q and deletion of 15q. *Am. J. Med. Genet.* **34:** 183–186.

Olson, M., Hood, L., Cantor, C., and Botstein, D. (1989). A common language for physical mapping of the human genome. *Science* **245:** 1434–1435.

Phillips, H. A., Harris, P., Richards, R. I., Sutherland, G. R., and Mulley, J. C. (1991a). Dinucleotide repeat polymorphisms at the *D16S164, D16S168* and *D16S186* loci at 16q21–q22.1. *Nucleic Acids Res.* **19:** 6964.

Phillips, H. A., Hyland, V. J., Holman, K., Callen, D. F., Richards, R. I., and Mulley, J. C. (1991b). Dinucleotide repeat polymorphism at *D16S287. Nucleic Acids Res.* **19:** 6664.

Pol, S., Bousquet-Lemercier, B., Pave-Preux, M., Pawlak, A., Nalpas, B., Berthelot, P., Hanoune, J., and Barouki, R. (1988). Nucleotide sequence and tissue distribution of the human mitochondrial aspartate aminotransferase mRNA. *Biochem. Biophys. Res. Commun.* **157:** 1309–1315.

Reeders, S. T., Hildebrand, C. E., and Sutherland, G. R. (1991). Human Gene Mapping 11. Report of the Committee on the genetic constitution of chromosome 16. *Cytogenet. Cell Genet.* **58:** 643–685.

Richards, R. I., Holman, K., Lane, S., Sutherland, G. R., and Callen, D. F. (1991). Human chromosome 16 physical map: Mapping of somatic cell hybrids using multiplex PCR deletion analysis of sequence tagged sites. *Genomics* **10:** 1047–1052.

Samson, L., Derfler, B., Boosalis, M., and Call, K. (1991). Cloning and characterization of a 3-methyladenine DNA glycosylase cDNA from human cells whose gene maps to chromosome 16. *Proc. Natl. Acad. Sci. USA* **88:** 9127–9131.

Sato, T., Akiyama, F., Sakamoto, G., Kasumi, F., and Nakamura, Y. (1991). Accumulation of genetic alterations and progression of primary breast cancer. *Cancer Res.* **51:** 5794–5799.

Shen, Y., Holman, K., Thompson, A., Kozman, H., Callen, D. F., Sutherland, G. R., and Richards, R. I. (1991). Dinucleotide repeat polymorphism at the *D16S288* locus. *Nucleic Acids Res.* **19:** 5445.

Smith, C. L., and Cantor, C. R. (1987). Purification, specific fragmentation and separation of large DNA molecules. *In* "Methods in Enzymology" (R. Wu, ed.), Vol. 155, pp. 449–467, Academic Press, San Diego.

Stallings, R. L., Doggett, N. A., Callen, D. F., Apostolou, S., Chen, L. Z., Nancarrow, J. K., Whitmore, S. A., Harris, P., Mitchison, H., Breuning, M., Saris, J. J., Fickett, J., Cinkosky, M., Torney, D. C., Hilderbrand, C. E., and Moyzis, K. (1992a). Evaluation of a cosmid contig physical map of human chromosome 16. Submitted for publication.

Stallings, R. L., Doggett, N. A., Okumura, K., and Ward, D. C. (1992b). Chromosome 16-specific repetitive DNA sequences that map to chromosomal regions known to undergo breakage/rearrangement in leukemia cells. *Genomics* **13:** 332–338.

Stallings, R. L., Torney, D. C., Hildebrand, C. E., Longmire, J. L., Deaven, L. L., Jett, J. H., Doggett, N. A., and Moyzis, R. K. (1990). Physical mapping of human chromosomes by repetitive sequence fingerprinting. *Proc. Natl. Acad. Sci., USA* **87:** 6218–6222.

Sutherland, G. R., and Hecht, F. (1985). "Fragile Sites on Human Chromosomes," Oxford Univ. Press, New York.

Suthers, G. K., Mulley, J. C., Voelckel, M. A., Dahl, N., Vaisanen, M. L., Steinbach, P., Glass, I. A., Schwartz, C. E., van Oost, B. A., Thibodeau, S. N., Haites, N. E., Oostra, B. A., Gine, R., Carballo, M., Morris, C. P., Hopwood, J. J., and Sutherland, G. R. (1991). Genetic mapping of new DNA probes at Xq27 defines a strategy for DNA studies in the fragile X syndrome. *Am. J. Hum. Genet.* **48:** 460–467.

Thompson, A. D., Shen, Y., Holman, K., Sutherland, G. R., Callen, D. F., and Richards, R. I. (1992). Isolation and characterisation of $(AC)_n$ microsatellite genetic markers from human chromosome 16. *Genomics* **13:** 402–408.

Trask, B. J. (1991). Fluorescence *in situ* hybridization: Applications in cytogenetics and gene mapping. *Trends Genet.* **7:** 149–154.

Tsuda, H., Zhang, W., Shimosato, Y., Yokota, J., Terada, M., Sugimura, T., Miyamura, T., and Hirohashi, S. (1990). Allelic loss on chromosome 16 associated with progression of human hepatocellular carcinoma. *Proc. Natl. Acad. Sci. USA* **87:** 6791–6794.

van Tuinen, P., Rich, D. C., Summers, K. M., and Ledbetter, D. H. (1987). Regional mapping panel for human chromosome 17: Application to neurofibromatosis type 1. *Genomics* **1:** 374–381.

Weber, J. L., Kwitek, A. E., and May, P. E. (1990). Dinucleotide repeat polymorphisms at the *D16S260, D16S261, D16S265, D16S266* and *D16S267* loci. *Nucleic Acids Res.* **18:** 4034.

Wilkie, A. O. M., Higgs, D. R., Rack, K. A., Buckle, V. J., Spurr, N. K., Fischel-Ghodsian, N., Ceccherini, I., Brown, W. R. A., and Harris, P. C. (1991). Stable length polymorphism of up to 260 kb at the tip of the short arm of human chromosome 16. *Cell* **64:** 595–606.

Witt, D. R., Lew, S. P., and Mann, J. (1988). Heritable deletion of band 16q21 with normal phenotype: Relationship to late replicating DNA. *Am. J. Hum. Genet.* **43:** A127 (Abstract).

Wong, Z., Royle, N. J., and Jeffreys, A. J. (1990). A novel human DNA polymorphism resulting from transfer of DNA from chromosome 6 to chromosome 16. *Genomics* **7:** 222–234.

Am. J. Hum. Genet. 52:922–927, 1993

# Incidence and Origin of "Null" Alleles in the (AC)n Microsatellite Markers

David F. Callen, Andrew D. Thompson, Yang Shen, Hilary A. Phillips, Robert I. Richards, John C. Mulley, and Grant R. Sutherland

Department of Cytogenetics and Molecular Genetics, Women's and Children's Hospital, North Adelaide, South Australia

## Summary

**Twenty-three (AC)n repeat markers from chromosome 16 were typed in the parents of the 40 CEPH (Centre d'Étude du Polymorphisme Humain) families. Where parents were informative, the entire families were then typed. There were seven markers in which null alleles were demonstrated, as recognized by the apparent noninheritance, by a sib, of a parental allele. Four of these markers showed a null allele in a single sibship, while in the other three at least 30% of the CEPH sibships were shown to have a null allele segregating. One null allele was sequenced and shown to be the result of an 8-bp deletion occurring within the priming sequence for PCR amplification of the (AC)n repeats. In gene mapping or in application to diagnosis, the presence of a segregating null allele will not corrupt the linkage data but could result in loss of information. In isolated instances a segregating null allele may be interpreted as nonpaternity. The presence of a null allele may generate misleading data when individuals are haplotyped to determine the presence of linkage disequilibrium with a disease gene.**

## Introduction

Microsatellite markers which depend on the variability in the length of (AC)n repeats are powerful tools for the genetic analysis of human populations (Weber 1990). Since they can exhibit high heterozygosity and are PCR formatted, they are ideal "index markers" on human chromosomes, for the construction of genetic maps and for application to mapping and diagnosis by linkage in disease families.

Analysis of these markers is dependent on PCR using oligoprimers which flank the (AC)n repeat. Any mutation which is within the DNA sequence complementary to the oligoprimers may inhibit or completely prevent their binding, resulting in either reduced or complete loss of product. Where PCR amplification is entirely prevented, this will be evident as an absence of PCR product. These "null" alleles will not necessarily be recognized when there is a product from the other homologue, and this may lead to an underestimate of marker heterozygosity and to apparent incompatability of genotypes within a family. Null alleles have long been known for protein polymorphisms and, more recently, for VNTR markers (Chakraborty et al. 1992) and have been recognized, together with population subdivision, as a major factor in depression of observed heterozygosity, compared with that expected on the basis of Hardy-Weinberg equilibrium. This report presents a summary of the incidence of null alleles for a number of (AC)n markers on chromosome 16. The molecular basis of one of these null alleles has been determined.

Received August 25, 1992; revision received November 16, 1992.
Address for correspondence and reprints: Dr. David F. Callen, Department of Cytogenetics and Molecular Genetics, Women's and Children's Hospital, North Adelaide, S.A. 5006, Australia.

0002-9297/93/5204-

0002-9297/93/5205-0009$02.00

## Material and Methods

The (AC)n repeat markers from chromosome 16 which were examined for the presence of a null allele are listed in table 1. All (AC)n repeats listed were first typed in the parents of the 40 CEPH (Centre d'Étude

**Table 1**

**Chromosome 16 Microsatellite Markers Genotyped on the CEPH Families**

| Locus (probe) | Size of Forward Primer, Size of Reverse Primer (bp) | Size of Product (bp) | No. of Alleles | Heterozygosity (%) | | No. of CEPH Families | |
|---|---|---|---|---|---|---|---|
| | | | | Observed | Expected | Typed | With Null Allele |
| Short arm: | | | | | | | |
| *D16S67* (CRI-0391)[a] ..... | 37, 25 | 149–165 | 9 | 48 | 77 | 31 | 17 |
| *D16S79A* (66F3)[b] ..... | 26, 24 | 153–169 | 9 | 52 | 79 | 33 | 13 |
| *D16S94* (VK5)[c] ..... | 24, 26 | 82–90 | 5 | 58 | 51 | 33 | 0 |
| *D16S287* (16XE81)[d] ..... | 25, 25 | 203–227 | 10 | 78 | 78 | 40 | 1 |
| *D16S291* (AC2.5)[e] ..... | 25, 25 | 154–170 | 9 | 83 | 78 | 38 | 0 |
| *D16S292* (AC2.3)[e] ..... | 24, 24 | 180–198 | 10 | 78 | 73 | 39 | 0 |
| *D16S294* (AC1)[e] ..... | 25, 24 | 136–138 | 2 | 38 | 49 | 23 | 1 |
| *D16S295* (62F3)[f] ..... | 24, 25 | 110–124 | 8 | 67 | 66 | 35 | 0 |
| *D16S296* (62B4)[f] ..... | 28, 25 | 150–162 | 8 | 76 | 75 | 38 | 0 |
| *D16S297* (15H1H)[f] ..... | 27, 25(R1)[g] | 176–184 | 5 | 45 | 61 | 28 | 17 |
| | 27, 25(R2) | 232–244 | 7 | 81 | 74 | 28 | 0 |
| *D16S298* (3.12)[e] ..... | 28, 25 | 172–192 | 11 | 70 | 79 | 37 | 0 |
| *D16S299* (6.17)[e] ..... | 25, 25 | 126–140 | 8 | 66 | 72 | 36 | 0 |
| *D16S300* (AC1.1)[e] ..... | 24, 25 | 163–175 | 7 | 76 | 77 | 36 | 0 |
| *D16S319* (AC7.14)[h] ..... | 20, 20 | 149–159 | 7 | 49 | 52 | 29 | 0 |
| Long arm: | | | | | | | |
| *D16S164* (AC16.15)[i] ..... | 29, 21 | 169–179 | 5 | 38 | 41 | 24 | 0 |
| *D16S186* (AC16.101)[i] ..... | 25, 25 | 130–178 | 10 | 57 | 63 | 35 | 0 |
| *D16S261* (MFD24)[j] ..... | 20, 20 | 88–100 | 6 | 67 | 71 | 36 | 0 |
| *D16S265* (MFD23)[j] ..... | 24, 24 | 160–184 | 13 | 77 | 79 | 38 | 0 |
| *D16S289* (AC7.46)[e] ..... | 28, 20 | 156–176 | 8 | 76 | 77 | 36 | 0 |
| *D16S301* (AC6.21)[e] ..... | 27, 25 | 142–152 | 6 | 64 | 64 | 35 | 0 |
| *D16S303* (AC6.26)[e] ..... | 27, 25 | 101–115 | 6 | 36 | 41 | 25 | 0 |
| *D16S304* (AC1.14)[e] ..... | 26, 28 | 132–154 | 12 | 61 | 60 | 33 | 1 |
| *D16S305* (AC1.15)[e] ..... | 25, 25 | 172–200 | 14 | 80 | 82 | 39 | 1 |

[a] Reverse primer sequence = 5′CCACAGCTGGAAACTGGAGAATCAG3′; forward primer sequence = 5′GACTCTGTCTCAAAAAAAAA-AAAAATACAATATATGA3′.

[b] Reverse primer sequence = 5′CAAAGTCTTGGCAGGTATGGACAACT3′; forward primer sequence = 5′ATGAGCCACCAAGCCCG-GGCAGAA3′

[c] Details of primer sequences and PCR conditions are given by Aksentijevich et al. (submitted).

[d] Details of primer sequences and PCR conditions are given by Phillips et al. (1991*b*).

[e] Details of primer sequences and PCR conditions are given by Thompson et al. (1992).

[f] Details of primer sequences and PCR conditions are available from Genome Data Base, Baltimore.

[g] Two different reverse primers were used (see text).

[h] Details of primer sequences and PCR conditions are given by Shen et al. (1992).

[i] Details of primer sequences and PCR conditions are given by Phillips et al. (1991*a*).

[j] Details of primer sequences and PCR conditions are given by Weber et al. (1990).

du Polymorphisme Humain) families. Where parents were informative, the entire families were then typed. As described elsewhere (Thompson et al. 1992), the products of PCR reactions were electrophoresed on denaturing polyacrylamide gels to allow resolution of the lengths of (AC)n repeats.

Observed heterozygosity was determined from the proportion of the 80 CEPH parents who were heterozygous. The expected heterozygosity was calculated from the allele frequencies estimated from these CEPH parents (160 chromosomes).

Sequencing of PCR-amplified DNA was a modified version of Murray's (1989) method. PCR product was purified for sequencing by using the Magic™ PCR Preps DNA Purification System (Promega). PCR and PCR sequencing used the same reaction conditions and

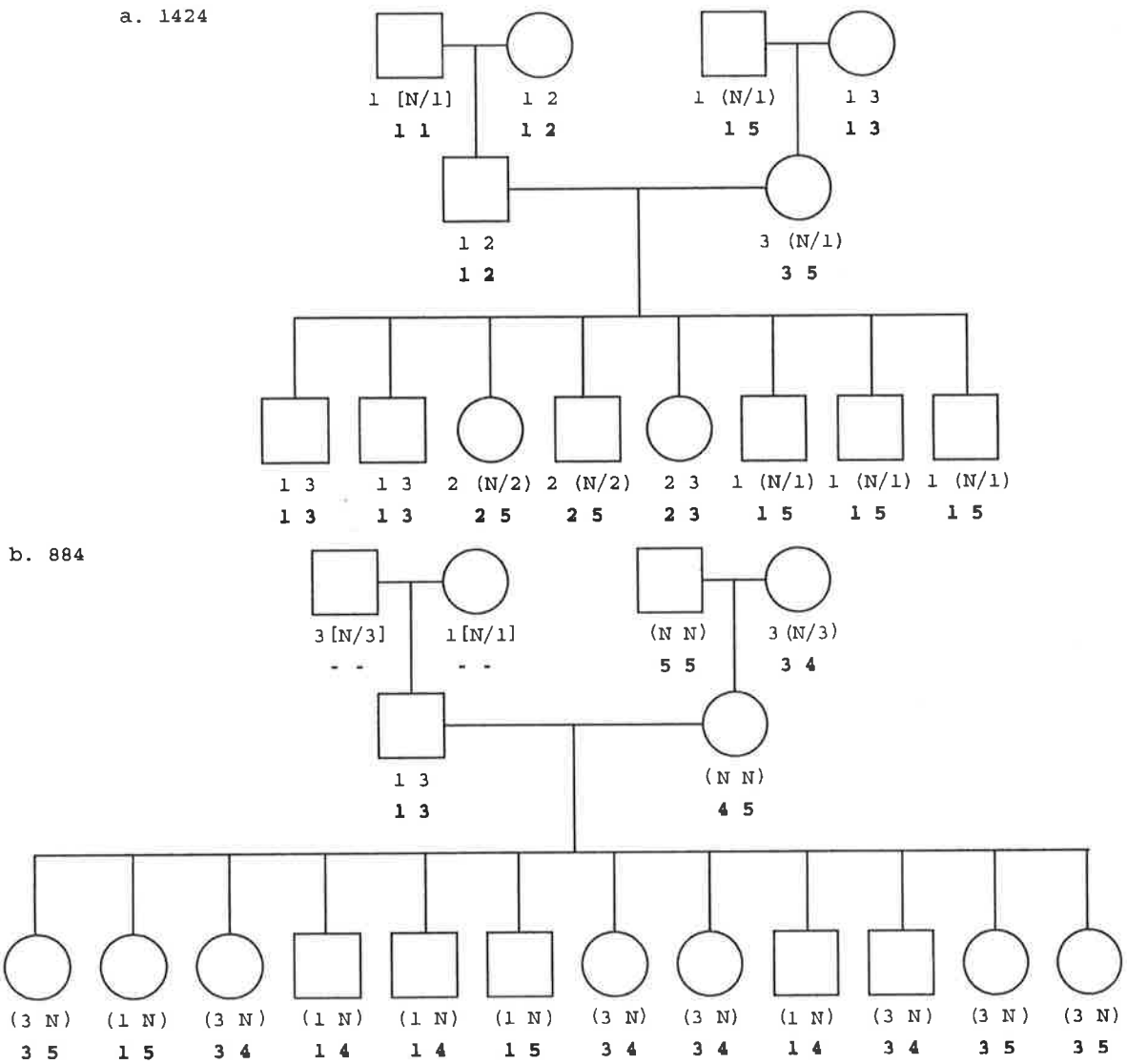



**Figure 1** CEPH pedigrees 1424 (*a*) and 884 (*b*) genotyped for *D16S297*. The upper row of genotypes corresponds to those obtained using the reverse primer R1. Inferred genotypes involving null alleles are indicated in parentheses, beside the marker phenotype. Genotypes which could potentially involve null alleles are indicated by square brackets beside the marker phenotype. The lower row of genotypes was scored using the R2 reverse primer which enabled detection of all null alleles. The alleles correspond to the following size classifications: 1, $(AC)_{15}$; 2, $(AC)_{13}$; 3, $(AC)_{12}$; 4, $(AC)_{15}\Delta 8$ bp; 5, $(AC)_{14}\Delta 8$ bp; and N, null.

step-cycle file as described elsewhere (Thompson et al. 1992), except that, for sequencing, dNTPs were at 7.5 μM, and $MgCl_2$ was at 4.5 mM. Purified product, primer (120 ng of 24mer), and 1.5 U *Taq* polymerase (Perkin Elmer Cetus) were divided into four 20-μl sequencing reactions, to each of which was added either 2.5 mM ddATP, 1.0 mM ddCTP, 0.5 mM ddGTP, or 5.0 mM ddTTP. Five microliters of each reaction was mixed with 4 μl formamide loading buffer and electrophoresed on a 6% wedge denaturing polyacrylamide gel, and the result was visualized by autoradiography.

## Results

Table 1 shows the 14 (AC)n repeats analyzed on the short arm of chromosome 16 and the 9 analyzed on the

long arm. The segregation of a null allele in a sibship was recognized by the apparent noninheritance, by a sib, of a parental allele (fig. 1). Of the 23 (AC)n repeats studied, 7 (30%) could be demonstrated to have null alleles. In four cases there was only a single CEPH sibship, of the 40 CEPH studied, where a null allele could be detected, while in the other three cases at least 30% of the CEPH sibships were shown to have a null allele segregating.

Where only a single CEPH sibship demonstrated the presence of a null allele, the number of sibs for whom a null allele was inferred to preserve Mendelian inheritance was four (*D16S287*), one (*D16S294*), four (*D16S304*), and five (*D16S305*). For each sibship the typing of the DNA samples was repeated, and the same results were obtained. The possibility of error due to incorrect DNA samples was eliminated, since the same CEPH sibship DNA samples have been used to determine the genotypes of all the (AC)n repeats listed in the table, and these data were all consistant with Mendelian inheritance.

There was no apparent relationship between the sequence or position of the oligoprimers flanking the (AC)n repeats and the presence of a null allele. For those (AC)n repeats with no demonstrated null allele, oligoprimers were constructed ranging from immediately adjacent to either side of the repeat (*D16S303*) to more than 20 bp either side of the repeat (*D16S296*). For those (AC)n repeats with a demonstrated null allele, the distance of the primers from the repeat was (presented as forward and reverse) 1 and 55 (*D16S67*), 2 and 69 (*D16S79A*), 14 and 104 (*D16S287*), 20 and 5 (*D16S294*), 32 and 53 (*D16S297*), 34 and 13 (*D16S304*), and 55 and 1 (*D16S305*).

The frequency of null alleles will be underestimated, since there may be no child within the pedigree who has the appropriate genotype to enable detection of a null allele. Not all 40 of the CEPH sibships were typed for each marker, with the families not typed being discarded as noninformative on the basis of parental genotypes. Since these apparently homozygous parents may in fact be heterozygous for a null allele, this will also underestimate the allele frequency.

Except for those markers with an appreciable frequency of null alleles, there is good agreement between expected and observed heterozygosity; hence population subdivision can be excluded as a contributor to the depression of heterozygosity observed for *D16S67*, *D16S79A*, and *D16S297*. Several CEPH sibships are known to be related, but this would have little effect on the calculation of expected heterozygosities. Where a null allele was present at an appreciable frequency, this was reflected in a lower observed heterozygosity, compared with that calculated from allele frequencies (table 1). For the (AC)n repeat at *D16S297*, 16 of the CEPH sibships were shown to be segregating a null allele. One of these sibships (sibship 884) had individuals homozygous for the null allele and therefore did not generate a PCR product (fig. 1*b*). This was investigated further by constructing an additional reverse oligoprimer—flanking the original reverse oligoprimer—to generate a larger product. Use of this reverse primer enabled (1) a PCR product to be detected for these null alleles and (2) two new alleles to be detected (table 1 and fig. 1). Typing of those CEPH sibships with suspected null alleles increased the observed heterozygosity from 45% to 81% (table 1), which is comparable to the expected frequency of 74%.

For the *D16S297* locus there were individuals who initially were homozygous for a null allele but who, with amplification using the additional reverse primer, were shown to be homozygous for one of the two new alleles. A homozygote for each of these new alleles was selected, and DNA was amplified by PCR using the additional reverse primer and was sequenced. For each individual there was an 8-bp deletion, and 4 bp within this deletion were part of a small direct repeat present in the undeleted sequence (fig. 2). This 8-bp deletion was located at the 3′ end of the sequence from which the first reverse primer was designed. This could either prevent primer extension or cause the reverse primer to fail to bind, resulting in absence of a PCR product, which would therefore be scored as a null allele. This deletion was associated with two alleles with different numbers of the (AC) repeat—i.e., alleles $(AC)_{14}\Delta 8$ bp and $(AC)_{15}\Delta 8$ bp—with frequencies of .03 and .22, respectively. The alleles which had dinucleotide-repeat number equivalent to these but which did not have the 8-bp deletion—i.e., alleles $(AC)_{14}$ and $(AC)_{15}$—had frequencies of .39 and .04, respectively.

## Discussion

This study of (AC)n microsatellite repeats on chromosome 16 demonstrates that null alleles can commonly occur. In the one case investigated in detail, the null allele was shown to be the result of an 8-bp deletion occurring within the sequence chosen as a priming sequence for PCR amplification of the (AC) repeat. This problem was overcome by synthesis of a new oli-

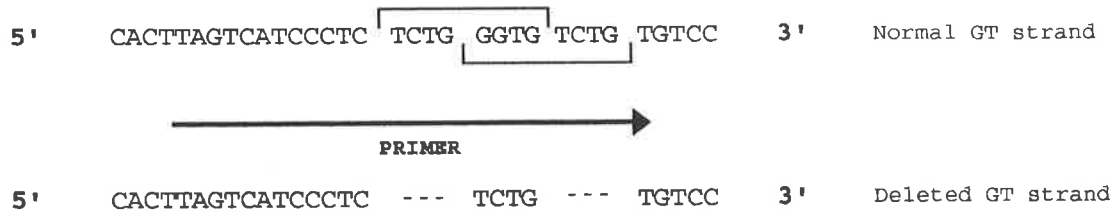



**Figure 2** Sequence of null alleles at *D16S297*, showing 8-bp deletion at site of reverse primer, i.e., R1. The two possible sites of the deletion on the intact strand are indicated in the upper sequence. The sequence used for the R1 primer is indicated on the lower sequence.

goprimer which did not include the site of this deletion. In the case of *D16S67* the forward primer includes the poly(A) sequence adjacent to an Alu repeat, and mutations in this sequence are likely to be responsible for the null allele. Examples of oligoprimers which fail to amplify alleles in some individuals have been previously noted for *D16S287* (Phillips et al. 1991*b*) and for *IL9* on chromosome 5 (Weber et al. 1991) and are likely to be due to a mechanism similar to that detailed in the present report. Alleles have been reported where there is inhibition of the PCR reaction, resulting in faint bands—e.g., *D16S186* (Phillips et al. 1991*a*) and *D14S34* (Weber et al. 1991). These cases are likely to be the result of polymorphisms within the binding site, which merely inhibits, rather than completely prevents, the amplification by PCR.

A null allele segregating in a single CEPH sibship was demonstrated for four of the (AC)n repeats on chromosome 16. For each of these (AC)n repeats, all other CEPH sibships typed were consistent with Mendelian inheritance. The same DNA samples were used for the determination of genotypes for all the loci in table 1. Therefore, it is unlikely that the genotypes interpreted as null alleles were due to technical factors. Further evidence for a null allele at *D16S287* has been described elsewhere (Phillips et al. 1991*b*). An additional pair of oligoprimers flanking this (AC) repeat were constructed and used to amplify the alleles in the family where segregation of a null allele was suspected. Mendelian inheritance was then observed, confirming the presence of a null allele.

Either in gene mapping or in application to diagnosis, the presence of a segregating null allele will not corrupt the linkage data. If undetected, a null allele will merely result in that individual being scored as a homozygote, resulting in loss of informativeness. The presence of a null allele at an appreciable frequency can be suspected (within a homogeneous population) when the observed heterozygosity is markedly less than the expected heterozygosity. The presence of an undetected null allele may generate misleading data when individuals are haplotyped to determine the presence of linkage disequilibrium of an (AC)n repeat with a disease gene.

Within a pedigree, vertical transmission of a null allele through apparent homozygotes can result in an individual's genotype being apparently inconsistent with classical Mendelian inheritance. In such cases an explanation of nonpaternity is often invoked, but, if the genotypes have been determined on the basis of PCR amplification, then the presence of a null allele should be considered as an alternative explanation. Synthesis of alternative oligoprimers should alleviate such problems.

Recently, Chakraborty et al. (1992) have suggested that, for forensic applications in DNA typing, PCR-based polymorphisms provide the ideal systems, since use of VNTRs can result in underestimation of heterozygotes, because of nondetection of small alleles on Southern blots. In view of the occurrence of null alleles in PCR-based systems, this recommendation should be approached with caution.

## Acknowledgments

This work was supported by U.S. Department of Energy contract DEFG0289ER60863.

## References

Aksentijevich I, Pras E, Gruberg L, Shen Y, Holman K, Helling S, Prosen L, et al. Refined physical mapping of the gene causing familial Mediterranean fever: a comparison of likelihood methods with homozygosity mapping (submitted)

Chakraborty R, de Andrade M, Daiger SP, Budowle B (1992) Apparent heterozygote deficiencies observed in DNA typing data and their implications in forensic applications. Ann Hum Genet 56:45–57

Murray V (1989) Improved double-stranded DNA sequenc-

ing using the linear polymerase chain reaction. Nucleic Acids Res 17:8889

Phillips HA, Harris P, Richards RI, Sutherland GR, Mulley JC (1991*a*) Dinucleotide repeat polymorphisms at the D16S164, D16S168 and D16S186 loci at 16q21-q22.1. Nucleic Acids Res 19:6964

Phillips HA, Hyland VJ, Holman K, Callen DF, Richards RI, Mulley JC (1991*b*) Dinucleotide repeat polymorphism at D16S287. Nucleic Acids Res 19:6664

Shen Y, Thompson AD, Holman K, Callen DF, Sutherland GR, Richards RI (1992) Four dinucleotide repeat polymorphisms on human chromosome 16 at D16S289, D16S318, D16S319 and D16S320. Hum Mol Genet 1:773

Thompson AD, Shen Y, Holman K, Sutherland GR, Callen DF, Richards RI (1992) Isolation and characterisation of (AC)n microsatellite genetic markers from human chromosome 16. Genomics 13:402–408

Weber JL (1990) Informativeness of human (dC-dA)n-(dG-dT)n polymorphisms. Genomics 7:524–530

Weber JL, Kwitek AE, May AE (1990) Dinucleotide repeat polymorphisms at the D16S260, D16S261, D16S265, D16S266 and D16S267 loci. Nucleic Acids Res 18:4034

Weber JL, Polymeropoulos MH, May PE, Kwitek AE, Xiao H, McPherson JD, Wasmuth JJ (1991) Mapping of human chromosome 5 microsatellite DNA polymorphisms. Genomics 11:695–700

Am. J. Hum. Genet. 53:451–461, 1993

# Refined Mapping of the Gene Causing Familial Mediterranean Fever, by Linkage and Homozygosity Studies

Ivona Aksentijevich,* Elon Pras,* Luis Gruberg,‡ Yang Shen,‖ Katherine Holman,‖ Sharon Helling,* Leandrea Prosen,* Grant R. Sutherland,‖ Robert I. Richards,‖ Mark Ramsburg,# Michael Dean,# Mordechai Pras,‡,§ Christopher I. Amos,† and Daniel L. Kastner*

*Arthritis and Rheumatism Branch and †Laboratory of Skin Biology, National Institute of Arthritis and Musculoskeletal and Skin Diseases, Bethesda; ‡Department of Medicine F and §Heller Institute for Medical Research, Sheba Medical Center, Tel-Hashomer, Israel; ‖Department of Cytogenetics and Molecular Genetics, Adelaide Children's Hospital, North Adelaide, South Australia; and #Laboratory of Viral Carcinogenesis, National Cancer Institute, Frederick Cancer Research and Development Center, Frederick, MD

## Summary

**Familial Mediterranean fever (FMF) is an autosomal recessive disease characterized by attacks of fever and serosal inflammation; the biochemical basis is unknown. We recently reported linkage of the gene causing FMF (designated "*MEF*") to two markers on chromosome 16p. To map *MEF* more precisely, we have now tested nine 16p markers. Two-point and multipoint linkage analysis, as well as a study of recombinant haplotypes, placed *MEF* between *D16S94* and *D16S80,* a genetic interval of about 9 cM. We also examined rates of homozygosity for markers in this region, among offspring of consanguineous marriages. For eight of nine markers, the rate of homozygosity among 26 affected inbred individuals was higher than that among their 20 unaffected sibs. Localizing *MEF* more precisely on the basis of homozygosity rates alone would be difficult, for two reasons: First, the high FMF carrier frequency increases the chance that inbred offspring could have the disease without being homozygous by descent at *MEF.* Second, several of the markers in this region are relatively nonpolymorphic, with a high rate of homozygosity, regardless of their chromosomal location.**

## Introduction

Familial Mediterranean fever (FMF) is an autosomal recessive disease characterized by recurrent, self-limited attacks of fever with sterile peritonitis, pleurisy, and/or synovitis. The disease is most common in non-Ashkenazi Jewish (Sohar et al. 1967), Armenian (Schwabe and Peters 1974), Turkish (Ozdemir and Sokmen 1969), and Middle Eastern Arab (Barakat et al. 1986) populations. Before the introduction of colchicine prophylaxis, a large percentage of Israeli and Turkish FMF patients developed systemic amyloidosis, inevitably leading to chronic renal failure and death or dialysis (Ozdemir and Sokmen 1969; Pras et al. 1982). Although the biochemical basis of this disease is unknown, the clinical manifestations of FMF suggest a lesion in a molecule important to the understanding of inflammation in general.

We have recently mapped the FMF susceptibility gene, designated by the gene symbol "*MEF*" (McAlpine et al. 1991), to the short arm of chromosome 16 in non-Ashkenazi Jewish families from Israel (Pras et al. 1992). Multipoint linkage analysis placed *MEF* centromeric to *D16S84,* with a maximal lod score greater than 19, but did not establish flanking markers for *MEF.* Shohat et al. (1992) have subsequently shown that the FMF susceptibility gene is also linked to chromosome 16p in Armenian families.

Our initial study also made use of the strategy of homozygosity mapping (Smith 1953; Lander and Botstein 1987). This approach relies on the fact that, in inbred families, children with rare recessive diseases usually inherit both copies of the disease gene, along with markers on adjacent chromosomal intervals, from a common ancestor. Most affected offspring of consanguineous marriages should therefore be homozy-

Received September 17, 1992; revision received February 18, 1993.

Address for correspondence and reprints: Daniel L. Kastner, M.D., Ph.D., National Institutes of Health, Building 6, Room 112, Bethesda, MD 20892.

gous by descent for DNA markers on either side of a recessive disease susceptibility gene. In the case of FMF, the gene frequency in our study population (approximately .045) is sufficiently high that some inbred affected individuals may not be homozygous by descent at the disease locus. Such individuals would not be expected to be homozygous at *MEF*-associated marker loci. Nevertheless, among 19 affected offspring of consanguineous marriages, we found a much higher rate of homozygosity at both chromosome 16 loci than we would have predicted on the basis of chance alone (Pras et al. 1992).

In the present study we have sought to define the boundaries of the *MEF* region of chromosome 16. To this end, we examined seven additional DNA polymorphisms and expanded our study panel to 31 non-Ashkenazi Jewish families. Regional localization was accomplished by classical likelihood methods and by the analysis of recombinant haplotypes. We also studied locus homozygosity among affected offspring of inbred marriages, in order to identify an area of overlap among the homozygous chromosomal regions defined for each inbred individual. This overlap zone should correspond to the *MEF* candidate region established by more traditional methods. In addition, we investigated the effects of varying disease gene frequency and marker heterozygosity on the applicability of this approach.

## Subjects and Methods

### *Families and DNA Samples*

Families were recruited from a clinic at the Sheba Medical Center, Tel-Hashomer, Israel; FMF was diagnosed according to established clinical criteria (Sohar et al. 1967). This project was approved by the Human Experimentation Committee at the Sheba Medical Center. The study included 31 pedigrees (200 individuals, 94 of whom were affected). Peripheral blood lymphocytes were immortalized with Epstein-Barr virus (Anderson and Gusella 1984). DNA was purified from cell lines by SDS–proteinase K digestion, extraction with phenol and chloroform, and ethanol precipitation (Davis et al. 1986).

### *Southern Analysis*

Southern blot analyses were carried out using standard techniques (Cavenee et al. 1984; Reed and Mann 1985). DNA probes were $^{32}$P-labeled by random priming (Oligo-labeling kit; Pharmacia) according to the manufacturer's instructions. Hybridizations were performed at 42°C overnight in the presence of 0.25 mg of human placental DNA/ml. Two 30-min high-stringency washes were performed with 0.1 × SSC, 0.1% SDS at 60°C. Films were exposed for 1–10 d at −70°C.

### *PCR Amplification*

The $(AC)_n$ marker *D16S283* was typed by PCR using published primers (Harris et al. 1991). One primer was end-labeled with $^{32}$P by using polynucleotide kinase (New England Biolabs). PCR was carried out in a 10-μl volume containing 50 ng of genomic DNA, with 13.4 ng of each unlabeled primer, 20 M dNTP, 0.08 ng of $^{32}$P-labeled primer, in 15 mmol $MgCl_2$ PCR buffer, with 1.2 U of *Taq*I polymerase (Perkin-Elmer). Thirty cycles were performed (94°C for 1 min, 57°C for 1 min, and 72°C for 30 s), followed by a final extension of 10 min at 72°C. Aliquots of amplified DNA were mixed with 1 vol of loading buffer, denatured at 94°C for 4 min, and electrophoresed on a standard 8% polyacrylamide gel. On every gel a control DNA sequence (Sequenase 2.0 kit; United States Biochemical) was run as a molecular-weight standard. Typing for *D16S291* was performed as described elsewhere (Thompson et al. 1992).

An $(AC)_n$ marker at *D16S94* was identified by screening restriction digests of lambda VK5 (Hyland et al. 1989) with $^{32}$P-labeled (Amersham Multiprime kit) poly-(AC.GT) (Pharmacia). A *Hae*III digest of lambda VK5 was subcloned into *Sma*I-cut M13mp18, and the $(AC)_n$-containing sequence was identified by hybridization to $^{32}$P-labeled poly-(AC.GT). Sequence analysis allowed the design of PCR primers VK5ACF 5′ GAGCCCACAGACTTCTTTCCTGGG 3′ and VK5ACR 5′ CCCACACCATGTACAGCCAGTAGGTA 3′ with which to type copy-number polymorphism in the $(AC)_n$ repeat. The parental generation (160 chromosomes) from the CEPH (Centre d'Étude du Polymorphisme Humain, Paris) pedigrees was genotyped to reveal the following allele frequencies: 82 bp, .01; 84 bp, .01; 86 bp, .62; 88 bp, .35; and 90 bp, .02. The expected heterozygosity was .51, and the observed heterozygosity was .58. For genotyping the FMF families, each gel contained the subcloned M13 (86-bp) allele as a size marker. PCR cycle times and reaction conditions were as described by Richards et al. (1991).

### *Linkage Analysis*

Lod scores were calculated using the LINKAGE package of programs (version 5.10) (Lathrop et al. 1984). We used a model in which FMF is inherited as an autosomal recessive trait with a 95% penetrance in males, 70% penetrance in females, and gene frequency of .045 (Pras et al. 1992). Allele frequencies were taken

## Table 1

**Pairwise Lod Scores between *MEF* and Chromosome 16p Markers**

| Locus | PIC[a] | $\theta$[b] .00 | .01 | .02 | .05 | .10 | .15 | Maximal Lod Score[c] | $\hat{\theta}_f$[d] | $\hat{\theta}_m$[d] |
|---|---|---|---|---|---|---|---|---|---|---|
| *HBA* (5'HVR/*Rsa*I) | .70 | −.79 | 10.47 | 12.61 | 14.62 | 14.42 | 12.90 | 15.70 | .01 | .11 |
| *D16S83* (EKMDA2.1/*Rsa*I) | .23 | 9.30 | 11.39 | 11.70 | 11.56 | 10.38 | 8.36 | 11.38 | .02 | .03 |
| *D16S84* (CMM65/*Taq*I) | .36 | 6.20 | 9.09 | 9.55 | 9.62 | 8.70 | 7.43 | 10.15 | .001 | .05 |
| *D16S291* (AC2.5, PCR marker) | .77 | 11.52 | 15.83 | 16.71 | 17.10 | 16.10 | 13.80 | 17.16 | .03 | .04 |
| *D16S283* (SM7, PCR marker) | .63 | 16.87 | 20.49 | 21.08 | 20.88 | 18.91 | 16.26 | 21.24 | .01 | .03 |
| *D16S94* (VK5, PCR marker) | .41 | 13.17 | 13.83 | 13.88 | 13.37 | 11.84 | 10.14 | 14.21 | .04 | .001 |
| *D16S80* (24-1/*Taq*I) | .31 | 6.55 | 9.77 | 10.38 | 10.69 | 9.88 | 8.60 | 11.25 | .001 | .07 |
| *D16S82* (41-1/*Sac*I) | .27 | 3.66 | 4.57 | 4.94 | 5.25 | 4.99 | 4.37 | 5.73 | .10 | .001 |
| *D16S37* (16/02/*Msp*I) | .37 | −19.65 | −11.82 | −9.14 | −5.35 | −2.64 | −1.36 | .00 | .50 | .50 |

NOTE.—Data were calculated from 31 non-Ashkenazi Jewish families by using a recessive-disease model with a 95% male penetrance and 70% female penetrance.

[a] Source: Williamson et al. (1991).

[b] Sex-averaged recombination frequencies.

[c] When sex-specific recombination frequencies are allowed to vary independently.

[d] Sex-specific recombination frequency for maximal lod score.

from published sources (Williamson et al. 1991). Recombination frequencies were transformed to map distances by Kosambi's (1944) formula.

## Results

### *Linkage Analysis*

Table 1 is a summary of sex-averaged and sex-specific two-point lod scores between *MEF* and nine markers mapped to chromosome 16p, tested in 31 non-Ashkenazi Jewish families. Figure 1 depicts the relative map positions of these loci. All but one of the markers demonstrated linkage to *MEF*. Only *D16S37*, which is closest to the centromere, was not linked. *D16S94*, an $(AC)_n$ marker with a PIC of .41, gave a maximal sex-averaged lod score of 13.88 at $\theta = .02$. The more informative *D16S283* had a higher maximal sex-averaged lod score, 21.20, at $\theta = .03$. For six of the eight markers showing linkage, sex-specific recombination frequencies were consistent with the excess in male recombination observed near the telomere of chromosome 16p (Breuning et al. 1990; Julier et al. 1990; Keith et al. 1990; Reeders et al. 1991).

In order to place *MEF* on the map, relative to other markers, we performed multipoint linkage analysis. For highly polymorphic markers, we were only able to use two fixed marker loci, because of the prohibitive computational times required for multipoint analysis in consanguineous families with several unknown typings. We chose *D16S80* and *D16S283* as fixed markers, on the basis of their map positions and informativeness. Using the ILINK computer program, we estimated the

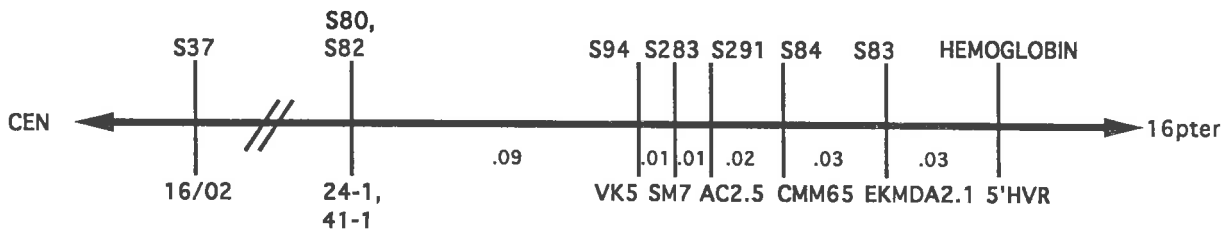



**Figure 1** Order of polymorphic sites within the *MEF* candidate region. Locus names and corresponding plasmid or microsatellite names are shown only for polymorphisms used in the present study. Sex-averaged recombination frequencies are taken from published sources (Breuning et al. 1990; Julier et al. 1990; Keith et al. 1990; Reeders et al. 1991) or are estimated from the present study.

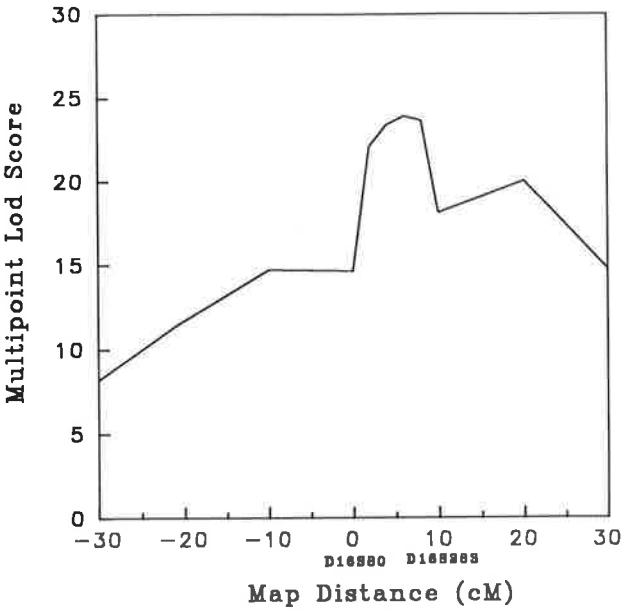



**Figure 2** Multipoint linkage analysis between *MEF* and two chromosome 16p markers, for 31 families. *D16S80* was arbitrarily assigned position 0, and *D16S283* was placed at 10 cM, on the basis of ILINK calculations with our panel of families. Multipoint lod scores were calculated every 2 cM within the interval from *D16S80* to *D16S283*. The maximal multipoint lod score was obtained 4 cM centromeric to *D16S283*.

recombination frequency between *D16S283* and *D16S80* in our families as .10, with no difference between sexes. The maximal multipoint lod score of 23.89 placed *MEF* between these two loci, about 4 cM centromeric to *D16S283* (fig. 2.) In light of the fact that *D16S94* is only about 200 kb centromeric to *D16S283* (Harris et al. 1991; Germino et al. 1992), *MEF* probably is in the interval between *D16S94* and *D16S80*.

### Recombinant Families

Figure 3 is a family-by-family summary of recombinants between *MEF* and chromosome 16 markers. Recombinant families are designated by blackened circles, and nonrecombinant families are designated by unblackened circles. All of the families except families 2 and 14 showed linkage to one or more loci. The latter two families were either recombinant or uninformative at all of the loci tested thus far. The true status of these families relative to chromosome 16 is unclear, since each family was uninformative for five of the nine markers in our panel.

A maternal crossover in family 15, present at *D16S94* but not at *D16S82*, placed *MEF* centromeric to *D16S94*. Families 33, 38, and 72 were not informative for *D16S94*, but in each of these families there were crossovers at *D16S283* that were not present at *D16S80* (families 33 and 38) or *D16S37* (family 72). The analysis of recombinant families was therefore consistent with the multipoint data placing *MEF* centromeric to *D16S94*.

A paternal crossover in family 27, present at *D16S80* but not at *D16S94*, placed *MEF* telomeric to *D16S80*. Moreover, in families 13 and 25 there were crossovers at *D16S80* that were not present at *D16S283*. When the centromeric and telomeric recombinants were considered together, the most likely position for *MEF* was between *D16S80* and *D16S94*.

### Homozygosity Mapping

Our panel of 31 non-Ashkenazi Jewish families included 12 consanguineous marriages, depicted in figure 4. There were seven first-cousin marriages, with a total of 18 affected offspring, and there were two uncle-niece marriages, with 3 affected offspring. In one other family (family 72) the offspring of half-sibs married, and in two families the exact relationship of intermarriage was unknown. Altogether, the panel included 26 affected offspring of inbred marriages.

Figure 5 shows whether each of these individuals was homozygous at the nine loci that we studied. Individuals 25-06, 27-11, and 38-08 were heterozygous at *D16S80* but homozygous at *D16S94*. Individuals 2-07 and 72-05 were heterozygous at *D16S94* but homozygous at *D16S80* and *D16S82*. There was no locus for which all affected individuals were homozygous.

We also evaluated whether the percentage of homozygous individuals increased as we moved closer to the *MEF* region (table 2). The left side of the table shows the fraction of homozygous individuals among the 26 affected and 20 unaffected offspring in our consanguineous families. For every locus except *D16S37* (which was not linked to *MEF* by two-point linkage analysis), the percentage homozygous among affected persons was greater than the percentage homozygous among unaffected siblings. For certain markers (*HBA, D16S83, D16S291,* and *D16S283*), this difference was quite substantial. For other markers (*D16S84, D16S94, D16S80,* and *D16S82*), the difference was not as great, owing largely to a high rate of homozygosity in the unaffected sibs.

The right side of table 2 shows the rate of homozygosity predicted for each marker, regardless of linkage to *MEF* and solely on the basis of chance, for various degrees of inbreeding. These values are close to the

| Family | 16/02 (S37) | 41-1 (S82) | 24-1 (S80) | VK5 (S94) | SM7 (S283) | 16AC2.5 (S291) | CMM65 (S84) | EKMDA (S83) | 5'HVR (HBA) |
|---|---|---|---|---|---|---|---|---|---|
| 2 | ● | ? | ? | ● | ? | ? | ? | ● | ● |
| 3 | O | O | O | O | O | O | ? | ? | O |
| 4 | ● | ? | O | O | O | O | O | O | O |
| 6 | O | O | O | O | O | O | ? | O | O |
| 7 | ? | ? | O | O | O | O | O | O | O |
| 8 | O | ? | ? | O | O | O | ? | O | O |
| 10 | ● | ? | ? | ? | O | O | O | ? | ● |
| 11 | ● | ? | ? | O | O | O | O | O | O |
| 13 | ? | ? | ● | ? | O | O | O | O | O |
| 14 | ? | ? | ? | ? | ● | ● | ? | ● | ● |
| 15 | ? | O | ? | ● | ● | ● | ● | ? | ? |
| 16 | ? | ? | O | O | O | O | O | ? | O |
| 17 | ? | ? | O | O | O | ? | O | O | O |
| 21 | ● | ● | ? | O | O | O | O | O | ? |
| 23 | ? | ? | O | O | O | O | ? | ● | ● |
| 25 | ● | ? | ● | ? | O | ? | O | O | O |
| 27 | ? | ? | ● | O | O | O | ? | ? | O |
| 30 | ? | O | ? | O | O | O | O | O | O |
| 31 | ? | ? | ? | ? | O | O | O | O | O |
| 33 | ● | O | O | ? | ● | ● | ● | ? | ● |
| 35 | ? | ? | ? | O | O | O | ? | O | ? |
| 38 | O | ? | O | ? | ● | ● | ● | ? | ● |
| 39 | ● | O | O | O | O | O | O | ? | O |
| 40/62 | ? | ● | O | ? | ? | ? | O | ? | O |
| 42 | O | ? | O | O | ? | O | O | O | O |
| 46 | ? | O | O | O | O | O | ? | ? | ? |
| 47 | ? | O | O | ? | O | ? | O | ? | ? |
| 49 | ● | ? | O | ? | ? | O | O | O | O |
| 60 | ? | ? | ? | O | O | O | ? | ? | O |
| 72 | O | ? | ? | ? | ● | ● | ● | ? | ● |
| 75 | ? | O | ? | O | O | O | O | O | ● |

**● = Recombination; O = No recombination; ? = uninformative or not done**

**Figure 3** Study of recombinants among 31 FMF families. Markers are ordered as in fig. 1. Family 15 defined the telomeric limit (*D16S94*) for the *MEF* candidate region, while families 13, 25, and 27 defined the centromeric limit (*D16S80*). The centromeric crossover in pedigree 33 was maternal, and the telomeric crossover was paternal.

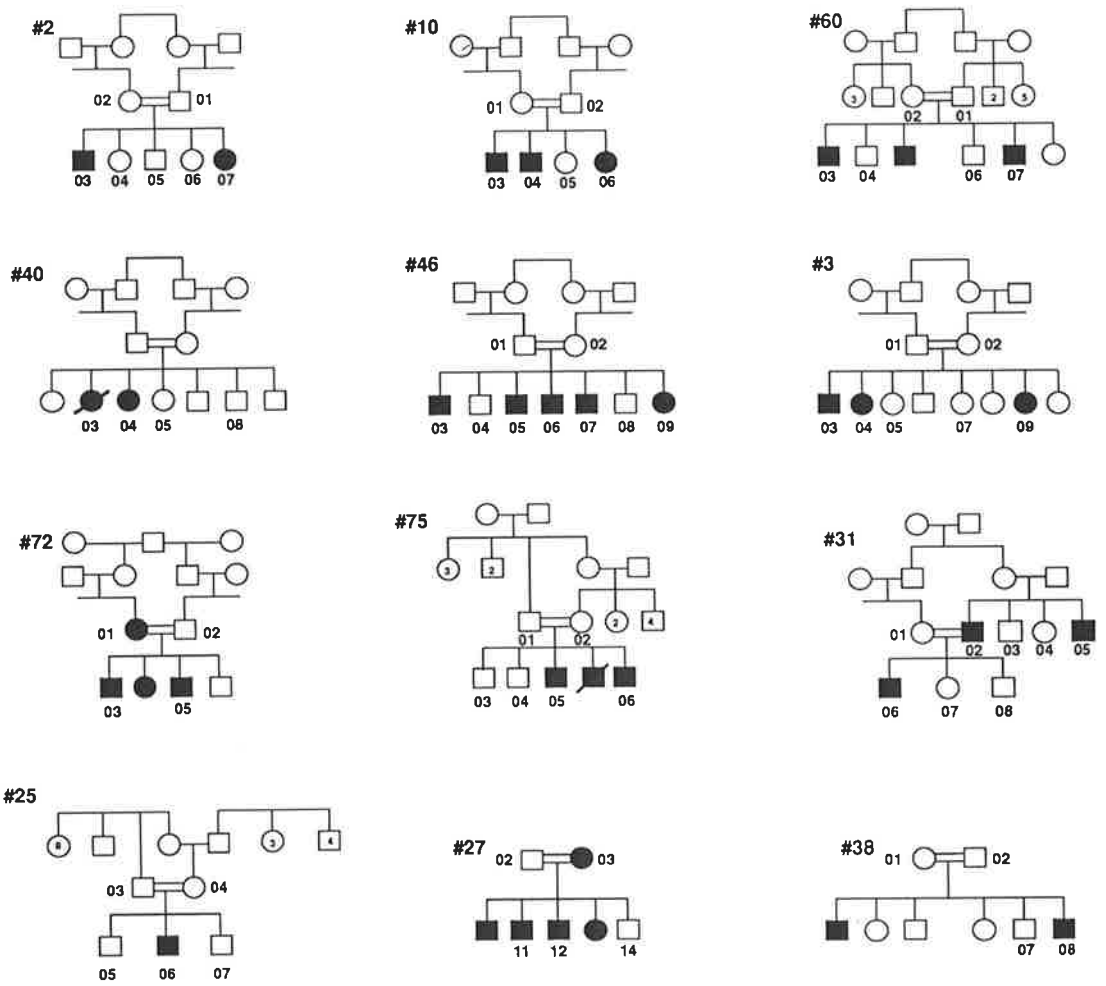



**Figure 4** Twelve consanguineous families used in the present study. DNA was available for each numbered individual.

percentages of homozygosity observed in the unaffected siblings in our inbred families. For relatively nonpolymorphic markers, high rates of homozygosity, even in unaffected individuals, are expected.

We derived a mathematical relationship (see the Appendix) to evaluate further the impact of disease gene frequency and marker polymorphism on the expected rates of homozygosity for markers linked to recessive disease genes. The value $P_H$ includes both the probability of observing marker homozygosity due to linkage between disease and marker loci and the probability of observing marker homozygosity by chance, independent of linkage between disease gene and marker locus.

Figure 6 shows $P_H$ as a function of $\theta$ for three disease gene frequencies and three marker heterozygosities, for affected offspring of first-cousin marriages. Panels *A–C* correspond, respectively, to the gene frequency for *MEF* in our population, the frequency of the cystic fibrosis allele in the United States (Boat et al. 1989), and the gene frequency for a rare recessive disease. Increasing disease gene frequency decreases $P_H$, because of an increased probability that carriers will marry into inbreeding loops. Increasing marker heterozygosity also decreases $P_H$, because this reduces the chances of observing homozygosity by state in the absence of homozygosity by descent.

## Discussion

The data in this paper confirm and substantially extend our previous report of linkage between *MEF* and two markers on the short arm of chromosome 16. We have added three RFLP markers and three PCR polymorphisms that are also linked to *MEF*. On the basis of

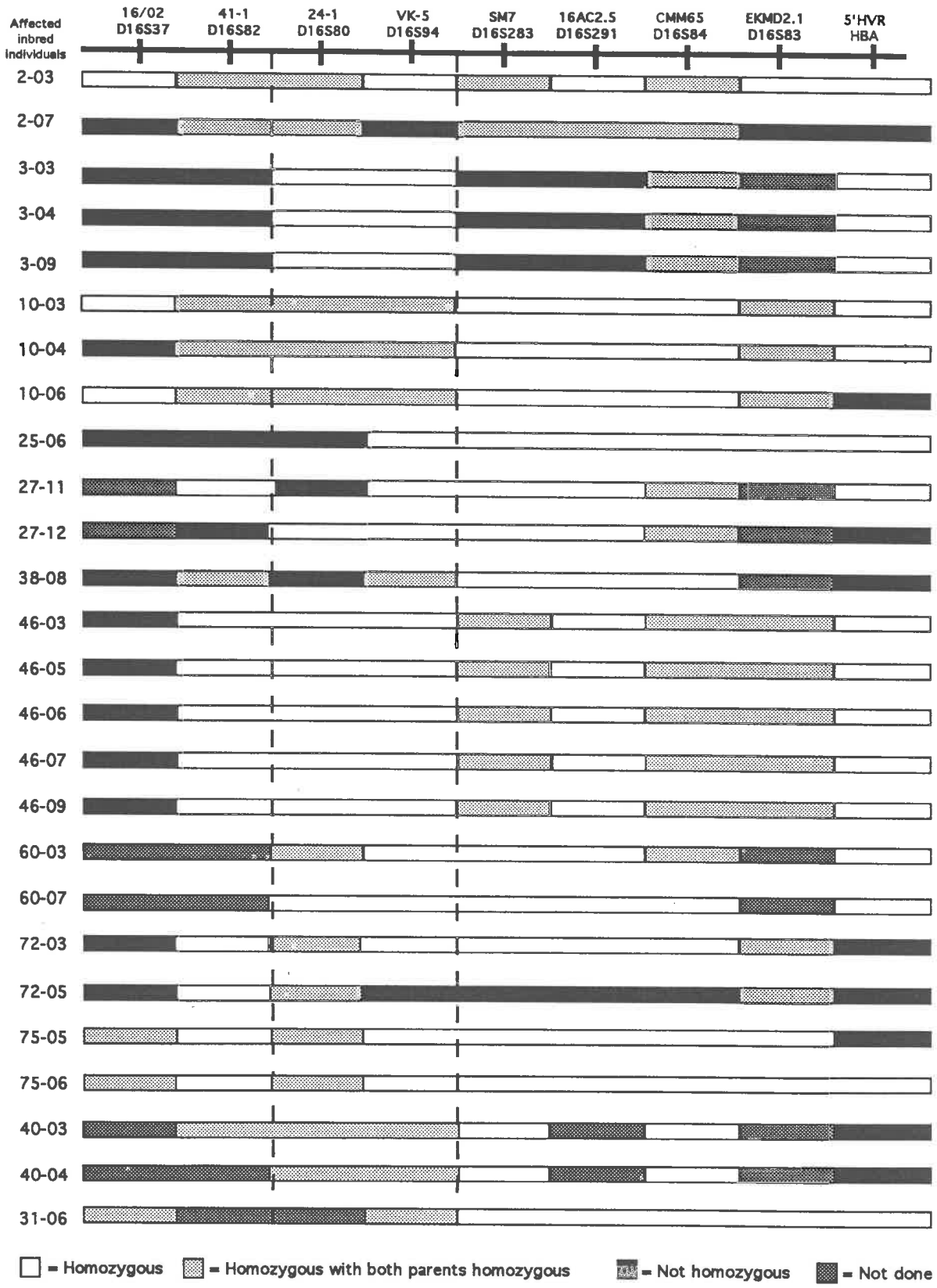



**Figure 5** Status of each individual for each locus, shown by shading of the appropriate block. These data are consistent with *MEF* mapping between *D16S94* and *D16S80*.

## Table 2

**Homozygosity Rates for Affected Offspring of Consanguineous Marriages**

| Marker | θ[a] | Observed Homozygosity | | | "Chance" Homozygosity[c] | | |
|---|---|---|---|---|---|---|---|
| | | Affected ($N$ = 26) | Unaffected ($N$ = 20) | $h$[b] | Uncle-Niece | First Cousins | Second Cousins |
| *HBA* (5'HVR) | .12 | .59 | .05 | .70 | .39 | .34 | .31 |
| *D16S83* (EKMDA) | .09 | .94 | .33 | .61 | .47 | .43 | .40 |
| *D16S84* (CMM65) | .06 | .96 | .70 | .47 | .59 | .56 | .54 |
| *D16S291* (16AC2.5) | .04 | .83 | .22 | .79 | .31 | .26 | .22 |
| *D16S283* (SM7) | .03 | .85 | .37 | .63 | .45 | .41 | .38 |
| *D16S94* (VK5) | .02 | .92 | .63 | .50 | .56 | .53 | .51 |
| *D16S80* (24-1) | .04 | .88 | .71 | .34 | .70 | .68 | .67 |
| *D16S82* (41-1) | .04 | .77 | .62 | .32 | .72 | .70 | .69 |
| *D16S37* (16/02) | .50 | .30 | .47 | .41 | .64 | .62 | .60 |

[a] Sex-averaged recombination between marker and *MEF*, based on the present study and published maps (Breuning et al. 1990; Julier et al. 1990; Keith et al. 1990; Reeders et al. 1991).

[b] Heterozygosity.

[c] Homozygosity expected for a random probe with heterozygosity $h$, for children of an inbred marriage, is $1-(1-F)h$, where $F$ is the coefficient of inbreeding (Hartl 1988).

linkage analysis and the study of recombinant families, we place *MEF* in the interval between *D16S94* and *D16S80*. None of the published maps of this region estimates this genetic interval (Breuning et al. 1990; Julier et al. 1990; Keith et al. 1990; Reeders et al. 1991). Our family data for 200 individuals indicate that the genetic distance between *D16S283* (which is very close to *D16S94*, both physically and genetically) and *D16S80* is about 10 cM, with $\theta_m = \theta_f$.

We have also typed 505 individuals from the CEPH reference pedigrees for *D16S283* and *D16S80* and have found the sex-averaged distance between these markers to be 9.6 cM. On the basis of the radiation hybrid data of Ceccherini et al. (1992) this interval is about 2.6 Mb, but the accuracy of such estimates remains uncertain. There are no known candidate genes that map to this region.

Among the 31 non-Ashkenazi families that we stud-

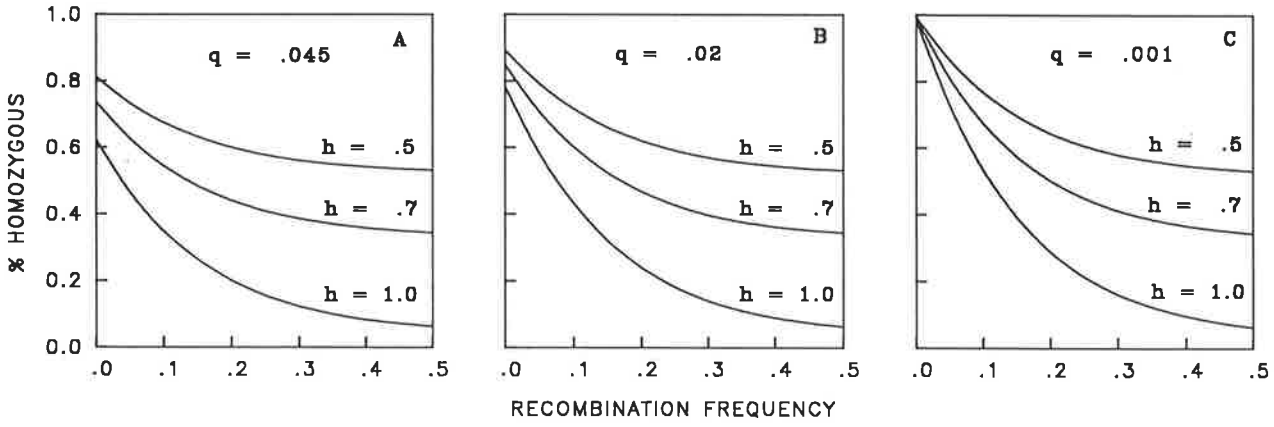



**Figure 6** Expected rate of homozygosity (i.e., $P_H$) plotted against recombination frequency (i.e., θ) in the affected offspring of first-cousin marriages. Each curve is calculated for a given disease allele frequency ($q$) and marker heterozygosity ($h$), according to the formula in the Appendix. $P_H$ includes both the probability of observing marker homozygosity due to linkage between disease and marker loci and the probability of observing marker homozygosity independent of linkage between disease gene and marker locus. This figure shows that homozygosity mapping is most reliable for fine localization for $q \ll 1$, $h \approx 1$, and θ approaching 0.

ied here, 12 were consanguineous. In those families where the relationship between parents was known, inbreeding loops were taken into account when likelihoods were calculated with the LINKAGE program. In many cases, this profoundly increased the information available for detecting linkage. All individuals in the inbreeding loops contribute informative meioses if a rare disease and a rare marker are segregating in a family. For instance, inclusion of the inbreeding loop connecting the uncle-niece marriage in family 25 led to a nearly fourfold increase in the lod score for *D16S283* at $\theta = 0$.

For relatively rare recessive diseases, markers in the chromosomal region encompassing the disease gene should be homozygous by descent in affected inbred individuals (Smith 1953; Lander and Botstein 1987). Hence, it may be possible to map a recessive disease gene with a panel of such individuals by searching for the area of overlap among homozygous chromosomal intervals. This approach was recently used to localize Werner syndrome to a 10-cM interval on chromosome 8 (Goto et al. 1992). Given the high frequency of the FMF susceptibility gene in our study population, we were unsure how useful this approach might be for us. However, in our initial study reporting linkage to chromosome 16, we observed very high rates of homozygosity for *HBA* and *D16S84* (Pras et al. 1992).

The data presented in figure 5 and table 2 confirm the utility of this strategy to demonstrate linkage to the short arm of chromosome 16. Rates of homozygosity were higher for affected inbreds than for their unaffected siblings, at eight of the nine markers that we examined. The data were particularly striking for the more polymorphic markers—*HBA, D16S83, D16S291,* and *D16S283*. The low rates of homozygosity at these markers for unaffected siblings argue against the possibility that locus homozygosity in affected individuals might be an inbreeding artifact independent of linkage to *MEF*.

However, as illustrated in figure 6, we encountered two limitations in the analysis of marker homozygosity alone for the fine localization of this gene. First, because of the high frequency of the FMF susceptibility gene in our study population, a substantial number of affected individuals need not be homozygous by descent. In light of a gene frequency of .045, the probability $\alpha$ that the affected child of an inbred marriage would be homozygous by descent at *MEF* is .76, .60, and .26 for uncle-niece, first-cousin, and second-cousin marriages, respectively (see the Appendix). Therefore, there may be no marker locus for which all of our inbred affected persons can be expected to be homozygous by descent. The second problem is that some of the relevant 16p markers are not very polymorphic and therefore have a high rate of homozygosity, especially in consanguineous families. This, coupled with the first problem, may give rise to a substantial percentage of affected inbred individuals who are homozygous for marker loci, regardless of the latters' map position.

For FMF, analysis of homozygous chromosome 16p intervals was consistent with the regional localization deduced from linkage studies and the analysis of recombinants. However, for the reasons discussed above, it would not have been possible, solely on the basis of marker homozygosity for our set of affected inbred individuals, to define the FMF interval as *D16S94–D16S80*. As has been the case for other disease genes, recombinant families will take on an increasingly important role as the FMF interval is narrowed further. Nevertheless, our experience with homozygosity mapping provides important insights, since it is one of the first practical applications of this strategy in human disease gene mapping.

## Acknowledgments

We thank Dr. Sherri Bale for stimulating discussions, and we thank Drs. Alisa Goldstein, Lynn R. Goldin, and Michael McDermott for reviewing the manuscript. We also thank Drs. Nathan Fischel-Ghodsian, Mordechai Shohat, and Jerome I. Rotter for sharing their data with us before publication. This work was supported in part by U.S. Department of Energy contract DEFG0289ER60863 and the Australian National Health and Medical Research Council. Computing time and staff support at the Advanced Scientific Computing Laboratory of the Frederick Cancer Research and Development Center were provided by the National Cancer Institute.

## Appendix

The probability of homozygosity in affected inbred individuals, for marker loci, is given by $P_H = \alpha\beta + \alpha(1-\beta)(1-h) + (1-\alpha)[1-(1-F)h]$, where $\alpha$ = probability that an affected child of the inbred marriage is homozygous by descent at the disease locus; $\beta$ = probability of homozygosity by descent at the marker locus, given homozygosity by descent at the disease locus; $F$ = coefficient of inbreeding; and $h$ = heterozygosity of the marker locus in the general population. The probability $\alpha$ can be calculated by $\alpha = Fq/[Fq+(1-F)q^2]$, where $q$ is the disease gene frequency (Lander and Botstein 1987). On the basis of the analysis of Smith (1953),

it can be shown that, for first-cousin marriages, $\beta = (1-\theta)^6 + \theta^2(1-\theta)^4 + \theta^2(1-\theta)^2/2$, where $\theta$ is the recombination frequency between disease locus and marker locus.

The first term of the equation for $P_H$ is the probability that the affected child is homozygous by descent at both the disease locus and the marker locus. The second term is the probability of an affected child being homozygous by descent at the disease locus but homozygous at the marker locus, by inheriting the same marker allele from different ancestors. The third term is the probability of an affected child being homozygous at the marker locus, given that he or she is not homozygous by descent at the disease locus. Thus, $P_H$ includes both the probability that an affected child is homozygous at the marker locus because of homozygosity by descent at a linked disease locus (first term) and the probability of homozygosity at the marker locus regardless of linkage to the disease locus (second and third terms).

## References


Anderson MA, Gusella JF (1984) Use of cyclosporin A in establishing Epstein-Barr virus-transformed human lymphoblastoid cell lines. In Vitro 20:856–858

Barakat MH, Karnik AM, Majeed HWA, el-Sobki NI, Fenech FF (1986) Familial Mediterranean fever (recurrent hereditary polyserositis) in Arabs: a study of 175 patients and review of the literature. Q J Med 60:837–847

Boat TF, Welsh MJ, Beaudet AL (1989) Cystic fibrosis. In: Scriver CR, Beaudet AL, Sly WS, Valle D (eds) The metabolic basis of inherited disease, 6th ed. McGraw-Hill, New York, pp 2649–2680

Breuning MH, Snijdewint FGM, Brunner H, Verwest A, Ijdo JW, Saris JJ, Dauwerse JG, et al (1990) Map of 16 polymorphic loci on the short arm of chromosome 16 close to the polycystic kidney disease gene (PKD1). J Med Genet 27: 603–613

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

Ceccherini I, Romeo G, Lawrence S, Breuning MH, Harris PC, Himmelbauer H, Frischauf AM, et al (1992) Construction of a map of chromosome 16 by using radiation hybrids. Proc Natl Acad Sci USA 89:104–108

Davis LG, Dibner MD, Battey JF (1986) Basic methods in molecular biology. Elsevier, New York

Germino GG, Weinstat-Saslow D, Himmelbauer H, Gillespie GAJ, Somlo S, Wirth B, Barton N, et al (1992) The gene for autosomal dominant polycystic kidney disease lies in a 750-kb CpG-rich region. Genomics 13:144–151

Goto M, Rubenstein M, Weber J, Woods K, Drayna D (1992) Genetic linkage of Werner's syndrome to five markers on chromosome 8. Nature 355:735–738

Harris PC, Thomas S, Ratcliffe PJ, Breuning MH, Cotto E, Lopez-Larrea C (1991) Rapid genetic analysis of families with polycystic kidney disease 1 by means of a microsatellite marker. Lancet 338:1484–1487

Hartl DL (1988) A primer of population genetics. Sinauer, Sunderland, MA

Hyland VJ, Fernandez KEW, Callen DF, MacKinnon RN, Baker E, Friend K, Sutherland GR (1989) Assignment of anonymous DNA probes to specific intervals of human chromosomes 16 and X. Hum Genet 83:61–66

Julier C, Nakamura Y, Lathrop M, O'Connell P, Leppert M, Mohandas T, Lalouel J-M, et al (1990) A primary map of 24 loci on human chromosome 16. Genomics 6:419–427

Keith TP, Green P, Reeders ST, Brown VA, Phipps P, Bricker A, Falls K, et al (1990) Genetic linkage map of 46 DNA markers on human chromosome 16. Proc Natl Acad Sci USA 87:5754–5758

Kosambi DD (1944) The estimation of map distances from recombination values. Ann Eugenics 12:172–175

Lander ES, Botstein D (1987) Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. Science 236:1567–1570

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

McAlpine PJ, Shows TB, Boucheix C, Huebner M, Anderson WA (1991) The 1991 catalog of mapped genes and report of the nomenclature committee. Cytogenet Cell Genet 58:5–102

Ozdemir AI, Sokmen C (1969) Familial Mediterranean fever among the Turkish people. Am J Gastroenterol 51:311–316

Pras E, Aksentijevich I, Gruberg L, Balow JE Jr, Prosen L, Dean M, Steinberg AD, et al (1992) Mapping of a gene causing familial Mediterranean fever to the short arm of chromosome 16. N Engl J Med 326:1509–1513

Pras M, Bronshpigel N, Zemer D, Gafni J (1982) Variable incidence of amyloidosis in familial Mediterranean fever among different ethnic groups. Johns Hopkins Med J 150:22–26

Reed KC, Mann DA (1985) Rapid transfer of DNA from agarose gels to nylon membranes. Nucleic Acids Res 13:7207–7221

Reeders ST, Hildebrand CE, Sutherland GR (1991) Report of the Committee on the Genetic Constitution of Chromosome 16. Cytogenet Cell Genet 58:643–685

Richards RI, Holman K, Kozman H, Kremer E, Lynch M, Pritchard M, Yu S, et al (1991) Fragile X syndrome: genetic localization by linkage mapping of two microsatellite repeats FRAXAC1 and FRAXAC2 which immediately flank the fragile site. J Med Genet 28:818–823

Schwabe AD, Peters RS (1974) Familial Mediterranean fever

in Armenians: analysis of 100 cases. Medicine (Baltimore) 53:453–462

Shohat M, Bu X, Shohat T, Fischel-Ghodsian N, Magal N, Nakamura Y, Schwabe AD, et al (1992) The gene for familial Mediterranean fever in both Armenians and non-Ashkenazi Jews is linked to the α-globin complex on 16p: evidence for locus homogeneity. Am J Hum Genet 51: 1349–1354

Smith CAB (1953) The detection of linkage in human genetics. J R Stat Soc [B] 15:153–184

Sohar E, Gafni J, Pras M, Heller H (1967) Familial Mediterranean fever: a survey of 470 cases and review of the literature. Am J Med 43:227–253

Thompson AD, Shen Y, Holman K, Sutherland GR, Callen DF, Richards RI (1992) Isolation and characterization of $(AC)_n$ microsatellite genetic markers from human chromosome 16. Genomics 13:402–408

Williamson R, Bowcock A, Kidd K, Pearson P, Schmidtke J, Ceverha P, Chipperfield M, et al (1991) Report of the DNA committee and catalogues of cloned and mapped genes, markers formatted for PCR and DNA polymorphisms. Cytogenet Cell Genet 58:1190–1832

*Am. J. Hum. Genet.* 53:644–651, 1993

# Familial Mediterranean Fever (FMF) in Moroccan Jews: Demonstration of a Founder Effect by Extended Haplotype Analysis

Ivona Aksentijevich,* Elon Pras,* Luis Gruberg,† Yang Shen,§ Katherine Holman,§ Sharon Helling,* Leandrea Prosen,* Grant R. Sutherland,§ Robert I. Richards,§ Michael Dean,‖ Mordechai Pras,‡ and Daniel L. Kastner*

*Arthritis and Rheumatism Branch, National Institute of Arthritis and Musculoskeletal and Skin Diseases, Bethesda; †Department of Medicine F and ‡Heller Institute for Medical Research, Sheba Medical Center, Tel-Hashomer, Israel; §Department of Cytogenetics and Molecular Genetics, Adelaide Children's Hospital, North Adelaide, South Australia; ‖Laboratory of Viral Carcinogenesis, National Cancer Institute, Frederick Cancer Research and Development Center, Frederick, MD

## Summary

**Familial Mediterranean fever (FMF) is an autosomal recessive disease causing attacks of fever and serositis. The FMF gene (designated "*MEF*") is on 16p, with the gene order *16cen–D16S80–MEF–D16S94–D16S283–D16S291–16pter*. Here we report the association of FMF susceptibility with alleles at *D16S94*, *D16S283*, and *D16S291* among 31 non-Ashkenazi Jewish families (14 Moroccan, 17 non-Moroccan). We observed highly significant associations at *D16S283* and *D16S291* among the Moroccan families. For the non-Moroccans, only the allelic association at *D16S94* approached statistical significance. Haplotype analysis showed that 18/25 Moroccan FMF chromosomes, versus 0/21 noncarrier chromosomes, bore a specific haplotype for *D16S94–D16S283–D16S291*. Among non-Moroccans this haplotype was present in 6/26 FMF chromosomes versus 1/28 controls. Both groups of families are largely descended from Jews who fled the Spanish Inquisition. The strong haplotype association seen among the Moroccans is most likely a founder effect, given the recent origin and genetic isolation of the Moroccan Jewish community. The lower haplotype frequency among non-Moroccan carriers may reflect differences both in history and in population genetics.**

## Introduction

Familial Mediterranean fever (FMF) is an autosomal recessive disease of unknown etiology, characterized by recurrent, self-limited attacks of fever with sterile peritonitis, pleurisy, and/or synovitis (Sohar et al. 1967). Patients may also develop systemic amyloidosis. Although the biochemical basis of this disease is unknown, the clinical manifestations of FMF suggest a lesion in a molecule important to the understanding of inflammation in general. We recently mapped the FMF susceptibility gene, designated by the gene symbol "*MEF*" (McAlpine et al. 1991), to chromosome 16p (Pras et al. 1992). Subsequent genetic linkage analysis of 31 non-Ashkenazi Jewish families from Israel placed *MEF* between *D16S80* and *D16S94*, a genetic interval of about 9 cM (Aksentijevich et al., in press). The mapping of *MEF* to chromosome 16p has been confirmed in Armenian families from California (Shohat et al. 1992).

Data associating particular alleles of marker loci with disease-causing mutations are sometimes helpful in the refined localization of a disease gene. Theoretically, there should be an inverse relationship between the degree of allelic association and physical or genetic distance (Hill and Robertson 1968). Allelic associations have focused interest on the appropriate physical map interval for the positional cloning of the genes causing cystic fibrosis (Kerem et al. 1989) and Huntington disease (Huntington's Disease Collaborative Research

Received March 5, 1993; revision received May 5, 1993.

Address for correspondence and reprints: Daniel L. Kastner, M.D., Ph.D., National Institutes of Health, Building 6, Room 112, Bethesda, MD 20892.

Group 1993). Linkage disequilibrium studies have also supported the mapping of the genes causing Friedreich ataxia (Fujita et al. 1990) and myotonic dystrophy (Harley et al. 1991).

In addition, when allelic or haplotype associations are uniquely found in specific subpopulations, this may provide evidence for a founder effect. Examples of this include choroideremia among Finns (Sankila et al. 1987), Friedreich ataxia among Louisiana Acadians (Sirugo et al. 1992) and southern Italians (Pandolfo et al. 1990), myotonic dystrophy among French Canadians (Harley et al. 1991), idiopathic torsion dystonia among Ashkenazi Jews (Ozelius et al. 1992), and cystic fibrosis among Hutterites (Fujiwara et al. 1989).

In the present study we examined allelic associations between FMF and chromosome 16p markers in a panel of 31 non-Ashkenazi Jewish families, in the hope that linkage disequilibrium would help further localize the disease gene within the *D16S80-D16S94* interval. Since a sizable subset of these families (14) had emigrated to Israel from Morocco, we looked for evidence of a founder effect in this separate group.

The possibility of a founder effect among Moroccan Jews appeared reasonable, given the history and genetics of this population. Most of the Moroccan Jewish community is relatively new, having been established by Jews who left Spain and Portugal around the time of the Spanish Inquisition (1492). There have been subsequent population "bottlenecks" due to religious persecution (Roth 1972). Genetically, Moroccan Jews have been relatively isolated. They have a much higher rate of Tay-Sachs disease than do other non-Ashkenazi populations, and mutations seen in Moroccan patients are different from those seen in Ashkenazi Jewish Tay-Sachs patients (Navon and Proia 1991). The Moroccan Jewish community also has a high rate of consanguinity (Sohar et al. 1961), which tends to obscure recombination and thereby extends the detection of founder chromosomes over more generations. The relatively recent origin, genetic isolation, and inbreeding in the Moroccan Jewish population all favor the identification of founder effects.

## Subjects and Methods

### *Families and DNA Samples*

Families were recruited from a clinic at the Sheba Medical Center, Tel-Hashomer, Israel. This project was approved by the Human Experimentation Committee at that institution. The study included 31 families with a total of 200 individuals (94 affected). Peripheral blood lymphocytes were immortalized with Epstein-Barr virus, and DNA was extracted by standard techniques. Among the 31 non-Ashkenazi families, 14 were of Moroccan origin (41 affected individuals); 9 were of Libyan, Tunisian, or Egyptian origin (33 affected individuals); 7 were of Iraqi or Kurdish origin (18 affected individuals); and 1 was of Greek origin (2 affected individuals). Our panel included 12 consanguineous families, with seven first-cousin, one half-first-cousin, and two uncle-niece marriages, as well as two marriages where the exact relationship of the partners was unknown.

### *DNA Markers*

The following probes were used for DNA markers flanking the *MEF* candidate region: *D16S84* (CMM65/*Taq*I), *D16S80* (24-1/*Taq*I), and *D16S82* (41-1/*Sac*I) as RFLP markers; and *D16S291* (16AC2.5), *D16S283* (SM7), and *D16S94* (VK5) as PCR markers. The latter three were typed as 8-allele, 11-allele, and 6-allele polymorphisms, respectively. In our study population we observed 44 haplotypes derived from these three loci.

DNA analysis by Southern blotting was performed by conventional procedures (Pras et al. 1992). PCR typing was done by using the published primers (Harris et al. 1991; Callen et al. 1992; Thompson et al. 1992), and reaction conditions were as described elsewhere (Aksentijevich et al., in press).

### *Statistical Analysis*

For each pedigree, we identified the parental alleles associated with disease susceptibility. For the 12 consanguineous marriages in the panel (see above), *MEF* and the associated chromosomal region usually appeared to be derived from a single ancestral chromosome. In these cases the disease-associated allele was counted only once. Control allele distributions were drawn from the noncarrier chromosomes in the same families. Haplotypes were derived so as to minimize recombinants.

To evaluate linkage disequilibrium, the Yule association coefficient was calculated according to the formula

$$|A| = |(ad-bc)/(ad+bc)|\ ,$$

where $a$ is the number of noncarrier chromosomes with allele A, $b$ is the number of FMF chromosomes with allele A, $c$ is the number of noncarrier chromosomes with allele B, and $d$ is the number of FMF chromo-

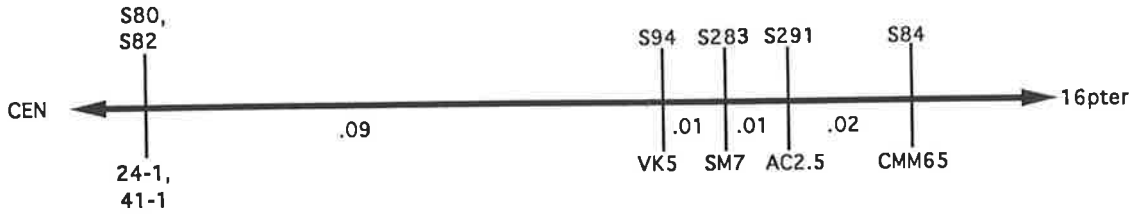



**Figure 1** Genetic map of the relevant region of chromosome 16p, showing the markers used in this study. Sex-averaged recombination frequencies are taken from published sources (Breuning et al. 1990; Reeders et al. 1991) and from our own data.

somes with allele B. The SD of the Yule coefficient was estimated according to Bishop et al. (1975). For multiallelic loci, the Yule coefficient was calculated for the allele that was overrepresented on FMF chromosomes, compared with all other alleles combined.

A $\chi^2$ test of the null hypothesis of no linkage disequilibrium was calculated for biallelic loci, with 1 df. The Yates correction was used where appropriate.

For multiallelic loci, two different $\chi^2$ statistics were calculated. In the "combined allele" method, the allele that was overrepresented among FMF chromosomes was defined as one allele, and all of the other alleles were combined to form a second allele. A standard $\chi^2$ statistic was then calculated, with 1 df. In the "multiallele" method, we used a $\chi^2$ statistic previously defined for multiallelic loci by Hill (1975) and Weir and Cockerham (1978). This is given by the formula

$$\chi^2 = N \sum_{i=1}^{m} \sum_{j=1}^{n} D_{ij}^2 / p_i q_j ,$$

where $N$ is the total number of gametes in the sample, $m$ is the number of alleles at the marker locus, $n$ (= 2) is the number of alleles at *MEF*, $p_i$ are the observed allele frequencies at the marker locus, $q_j$ are the observed frequencies of carrier and noncarrier chromosomes, and $D_{ij} = p_{ij} - p_i q_j$. This statistic has $(m-1) \times (n-1)$ df. Since $n = 2$, df $= m - 1$. When there are only 2 alleles, this $\chi^2$ formula yields the same result as a conventional $2 \times 2$ $\chi^2$ statistic.

## Results

### *Allelic Associations for All Non-Ashkenazi Families*

Our previous study of recombinants among 31 non-Ashkenazi Jewish families placed *MEF* in the interval between *D16S80* and *D16S94* (fig. 1; Aksentijevich et al., in press). We have subsequently analyzed allele frequencies in the same 31 families for the four telomeric and two centromeric loci shown in figure 1. Data for the highly polymorphic microsatellites at the telomeric end of the *MEF* interval are shown in table 1. When it is taken into account that we tested allelic associations at six different loci, the appropriate $P$ value for significance is .0083.

The 86-bp allele for *D16S94* was associated with disease susceptibility in 43 of 56 (77%) carrier chromosomes and was found in 25 of 48 (52%) noncarrier chromosomes. When allowance is made for multiple comparisons, this association approached (but did not reach) significance by both the combined allele and multiallele $\chi^2$ statistics. About 200 kb more telomeric (Harris et al. 1991; Germino et al. 1992), the 93-bp allele for *D16S283* was significantly associated with FMF susceptibility. At *D16S291*, still more telomeric, the 162-bp allele was associated with susceptibility (52% of carrier chromosomes vs. 23% of noncarrier chromosomes). This latter association was significant only by the combined allele method. There was no significant allelic association for *D16S84*, the most telomeric marker examined, nor were there significant associations for either centromeric marker (not shown).

### *Allelic Associations for Moroccan Families*

Families that had emigrated from Morocco to Israel constituted the largest single subpopulation of our sample (14 families). For *D16S94*, the percentages of carrier and noncarrier chromosomes bearing the 86-bp allele were nearly the same in the Moroccans as in the whole panel (table 2), and the Yule coefficients were nearly the same (0.47 vs. 0.51). The disease association for the 86-bp allele at *D16S94* did not approach statistical significance, mainly because of the smaller size of the Moroccan subset. For both *D16S283* and *D16S291*, allelic associations were much higher among the Moroccans than in the whole panel and were highly significant. These latter results are consistent with a founder effect in the Moroccan subpopulation. There were no

**Table 1**

**Allelic Associations for 31 Non-Ashkenazi Jewish Families**

| Marker and Allele | No. (%) of FMF Chromosomes | | $\chi^2$ [a] | | |
|---|---|---|---|---|---|
| | Carrier | Noncarrier | Combined Allele ($P$) | Multiallele ($P$) | $\|A\|$[b] |
| *D16S94:* | | | | | |
| 92 bp.... | 0 (0) | 1 (2) | | | |
| 90 bp.... | 1 (2) | 3 (6) | | | |
| 88 bp.... | 6 (11) | 15 (31) | | | |
| 86 bp[c]... | 43 (77) | 25 (52) | 5.63 (.018) | 12.07 (.034) | .51 ± .15 |
| 84 bp.... | 2 (4) | 3 (6) | | | |
| 82 bp.... | 4 (7) | 1 (2) | | | |
| *D16S283:* | | | | | |
| 103 bp.... | 2 (3) | 0 (0) | | | |
| 101 bp.... | 1 (2) | 0 (0) | | | |
| 99 bp.... | 2 (3) | 2 (3) | | | |
| 97 bp.... | 1 (2) | 1 (2) | | | |
| 95 bp.... | 0 (0) | 0 (0) | | | |
| 93 bp[c]... | 31 (53) | 9 (16) | 17.97 (.000023) | 28.82 (.0013) | .72 ± .11 |
| 91 bp.... | 17 (29) | 25 (43) | | | |
| 89 bp.... | 4 (7) | 16 (28) | | | |
| 87 bp.... | 0 (0) | 2 (3) | | | |
| 73 bp.... | 0 (0) | 1 (2) | | | |
| 65 bp.... | 0 (0) | 2 (3) | | | |
| *D16S291:* | | | | | |
| 168 bp.... | 0 (0) | 1 (2) | | | |
| 166 bp.... | 5 (10) | 8 (16) | | | |
| 164 bp.... | 13 (26) | 14 (29) | | | |
| 162 bp[c]... | 26 (52) | 11 (23) | 8.81 (.003) | 11.44 (.12) | .57 ± .15 |
| 160 bp.... | 1 (2) | 3 (6) | | | |
| 158 bp.... | 2 (4) | 4 (8) | | | |
| 156 bp.... | 3 (6) | 6 (12) | | | |
| 152 bp.... | 0 (0) | 1 (2) | | | |

[a] $\chi^2$ test of the null hypothesis of no linkage disequilibrium, calculated as described in Subjects and Methods. Significance levels are in parentheses. When allowance is made for the fact that we studied allelic associations for six loci, the appropriate $P$ value for significance is .0083.

[b] The Yule association coefficient ± SD.

[c] FMF-associated allele.

significant allelic associations for *D16S84*, nor for either centromeric locus.

### *Allelic Associations for Non-Moroccan Families*

Given the foregoing data, we hypothesized that the strong allelic associations in the Moroccan subpopulation might be the basis for the somewhat weaker associations in the full panel of families. Table 3 shows the data for just the 17 non-Moroccan families. Again using a critical $P$ value of .0083, we did not observe significant allelic associations among these non-Moroccan families. Only for the 86-bp allele at *D16S94* did the data approach significance. These findings are consistent with the hypothesis that, for *D16S283* and *D16S291*, allelic associations observed in the whole panel of families result from a founder effect among Moroccans.

### *Haplotype Associations*

Extended haplotypes for *D16S94–D16S283–D16S291* were constructed to identify possible founder haplotypes. Of a possible 528 haplotypes, 44 were observed in our study population. Among Moroccans, the 86-93-162 haplotype accounted for 18/25 (72%) of the carrier chromosomes, whereas this haplotype was not observed among 22 Moroccan noncarrier chromosomes. Among non-Moroccans, 6/26 (23%) carrier chromosomes bore this haplotype, versus 1/28 non-

**Table 2**
**Allelic Associations for 14 Moroccan Jewish Families**

| Marker and Allele | No. (%) of FMF Chromosomes | | $\chi^2$[a] | | |
|---|---|---|---|---|---|
| | Carrier | Noncarrier | Combined Allele (*P*) | Multiallele (*P*) | $\|A\|$[b] |
| *D16S94:* | | | | | |
| 92 bp. . . . | 0 (0) | 0 (0) | | | |
| 90 bp. . . . | 0 (0) | 2 (7) | | | |
| 88 bp. . . . | 5 (18) | 7 (26) | 3.00 (.083) | 4.33 (.50) | .47 ± .25 |
| 86 bp[c]. . . | 21 (78) | 15 (56) | | | |
| 84 bp. . . . | 1 (4) | 3 (11) | | | |
| 82 bp. . . . | 0 (0) | 0 (0) | | | |
| *D16S283:* | | | | | |
| 103 bp. . . . | 1 (4) | 0 (0) | | | |
| 101 bp. . . . | 0 (0) | 0 (0) | | | |
| 99 bp. . . . | 0 (0) | 0 (0) | | | |
| 97 bp. . . . | 1 (4) | 1 (4) | | | |
| 95 bp. . . . | 0 (0) | 0 (0) | | | |
| 93 bp[c]. . . | 22 (79) | 4 (15) | 19.93 (.000008) | 26.01 (.0037) | .91 ± .06 |
| 91 bp. . . . | 2 (7) | 11 (41) | | | |
| 89 bp. . . . | 2 (7) | 10 (37) | | | |
| 87 bp. . . . | 0 (0) | 1 (4) | | | |
| 73 bp. . . . | 0 (0) | 0 (0) | | | |
| 65 bp. . . . | 0 (0) | 0 (0) | | | |
| *D16S291:* | | | | | |
| 168 bp. . . . | 0 (0) | 1 (4) | | | |
| 166 bp. . . . | 1 (4) | 5 (21) | | | |
| 164 bp. . . . | 1 (4) | 7 (29) | | | |
| 162 bp[c]. . . | 20 (80) | 5 (21) | 17.15 (.000035) | 20.48 (.0046) | .88 ± .08 |
| 160 bp. . . . | 1 (4) | 0 (0) | | | |
| 158 bp. . . . | 1 (4) | 2 (8) | | | |
| 156 bp. . . . | 1 (4) | 3 (12) | | | |
| 152 bp. . . . | 0 (0) | 1 (4) | | | |

[a] $\chi^2$ test of the null hypothesis of no linkage disequilibrium, calculated as described in Subjects and Methods. Significance levels are in parentheses. When allowance is made for the fact that we studied allelic associations for six loci, the appropriate *P* value for significance is .0083.
[b] The Yule association coefficient ± SD.
[c] FMF-associated allele.

carrier chromosomes. The six non-Moroccan haplotype-positive carrier chromosomes included one from a Tunisian family, one from a Libyan family, two from separate Egyptian families, one from a Greek family, and one from an Iraqi family.

## Discussion

In 14 Moroccan Jewish families, we have found highly significant allelic and haplotype associations between FMF susceptibility and chromosome 16p markers. These markers—*D16S94*, *D16S283*, and *D16S291*—are tightly linked with one another at the telomeric end of the *MEF* interval.

Although recombinant events identify *D16S94* as the closest of the three markers to *MEF* (Aksentijevich et al., in press), there was no significant allelic association for this locus among the Moroccans. This is most likely because the 86-bp *D16S94* allele associated with FMF was also the most common allele in noncarrier chromosomes. In contrast, the 93-bp allele at *D16S283* and the 162-bp allele at *D16S291* were relatively uncommon in noncarriers. Tests of linkage disequilibrium are much more powerful when a disease gene is associated with a marker allele that is rare in the general population (Thompson et al. 1988).

Our data are consistent with the hypothesis that most FMF carrier chromosomes in the Moroccan Jew-

**Table 3**

**Allelic Associations for 17 Non-Moroccan Jewish Families**

| Marker and Allele | No. (%) of FMF Chromosomes | | $\chi^2$ [a] | | |
|---|---|---|---|---|---|
| | Carrier | Noncarrier | Combined Allele (*P*) | Multiallele (*P*) | \|*A*\| [b] |
| *D16S94:* | | | | | |
| 92 bp.... | 0 (0) | 1 (5) | | | |
| 90 bp.... | 1 (3) | 1 (5) | | | |
| 88 bp.... | 1 (3) | 8 (38) | | | |
| 86 bp[c]... | 22 (76) | 10 (48) | 4.22 (.04) | 12.79 (.025) | .55 ± .21 |
| 84 bp.... | 1 (3) | 0 (0) | | | |
| 82 bp.... | 4 (14) | 1 (5) | | | |
| *D16S283:* | | | | | |
| 103 bp.... | 1 (3) | 0 (0) | | | |
| 101 bp.... | 1 (3) | 0 (0) | | | |
| 99 bp.... | 2 (7) | 2 (6) | | | |
| 97 bp.... | 0 (0) | 0 (0) | | | |
| 95 bp.... | 0 (0) | 0 (0) | | | |
| 93 bp[c]... | 9 (30) | 5 (16) | 1.66 (.20) | 9.16 (.52) | .39 ± .27 |
| 91 bp.... | 15 (50) | 14 (45) | | | |
| 89 bp.... | 2 (7) | 6 (19) | | | |
| 87 bp.... | 0 (0) | 1 (3) | | | |
| 73 bp.... | 0 (0) | 1 (3) | | | |
| 65 bp.... | 0 (0) | 2 (6) | | | |
| *D16S291:* | | | | | |
| 168 bp.... | 0 (0) | 0 (0) | | | |
| 166 bp.... | 4 (16) | 3 (13) | | | |
| 164 bp.... | 12 (48) | 7 (29) | | | |
| 162 bp[c]... | 6 (24) | 6 (25) | .007 (.93) | 4.97 (.66) | .03 ± .33 |
| 160 bp.... | 0 (0) | 3 (13) | | | |
| 158 bp.... | 1 (4) | 2 (8) | | | |
| 156 bp.... | 2 (8) | 3 (13) | | | |
| 152 bp.... | 0 (0) | 0 (0) | | | |

[a] $\chi^2$ test of the null hypothesis of no linkage disequilibrium, calculated as described in Subjects and Methods. Significance levels are in parentheses. When allowance is made for the fact that we studied allelic associations for six loci, the appropriate *P* value for significance is .0083.

[b] The Yule association coefficient ± SD.

[c] FMF-associated allele.

ish population are derived from a common ancestor. The 86-93-162 haplotype for *D16S94–D16S283–D16S291* was present in 72% of 25 randomly chosen Moroccan carrier chromosomes. On the basis of allele frequencies in our Moroccan control chromosomes, and by assuming linkage equilibrium among these marker loci, we would have expected a haplotype frequency of .21 × .15 × .56, or 1.8%. On the basis of allele frequencies from other control populations (Harris et al. 1991; Thompson et al. 1992), the figure would be .25 × .04 × .62, or 0.6%. The 86-93-162 haplotype was therefore present in vast excess among Moroccan FMF carrier chromosomes.

To exclude the possibility that the families we studied were not representative of the broader Moroccan FMF population, we determined the city or village of origin for the parental generation of each of these families. We were able to trace the ancestry of these 14 families to 10 different communities, encompassing most of the habitable areas of Morocco. Moreover, none of the families reported any known common ancestry with any other family in the panel.

It is unclear whether the allelic and haplotype associations we observed at the telomeric end of the *MEF* interval actually place the disease gene closer to *D16S94* than to *D16S80*. Both of the centromeric markers we examined, *D16S80* and *D16S82*, are relatively nonpolymorphic, and in both cases the allele as-

sociated with FMF was the common allele in the general population (data not shown). Given the size of our sample, it would be difficult to establish significant allelic associations at these loci, and thus it is probably premature to use linkage disequilibrium data to infer that *MEF* is closer to *D16S94* than to *D16S80*.

The allelic and haplotype associations described here may also provide a basis for carrier screening for FMF. With a frequency of about 70% in Moroccan carriers, the 86-93-162 haplotype is not sensitive enough for diagnostic purposes at present. However, since we did not observe this haplotype at all among 22 Moroccan control chromosomes, the specificity may be very high, and this haplotype might be used to confirm diagnoses. Identification of new markers in the candidate region may broaden the applicability of this approach in selected groups.

## Acknowledgments

We thank Drs. Chris Amos, Sherri Bale, Lynn Goldin, and Michael McDermott for critically reading this manuscript, and we thank Drs. Nathan Fischel-Ghodsian, Mordechai Shohat, and Jerome I. Rotter for sharing their data with us before publication. This work was supported in part by the U.S. Department of Energy (contract DEFG0289ER60863) and the Australian National Health and Medical Research Council.

## References

Aksentijevich I, Pras E, Gruberg L, Shen Y, Holman K, Helling S, Prosen L, et al. Refined mapping of the gene causing familial Mediterranean fever, by linkage and homozygosity studies. Am J Hum Genet 53 (in press)

Bishop YNM, Fienberg S, Holland PW (1975) Discrete multivariate analysis: theory and practice. MIT Press, Cambridge and London

Breuning MH, Snijdewint FGM, Brunner H, Verwest A, Ijdo JW, Saris JJ, Dauwerse JG, et al (1990) Map of 16 polymorphic loci on the short arm of chromosome 16 close to the polycystic kidney disease gene (PKD1). J Med Genet 27:603–613

Callen DF, Doggett NA, Stallings RL, Chen LZ, Whitmore SA, Lane SA, Nancarrow JK, et al (1992) High-resolution cytogenetic-based physical map of human chromosome 16. Genomics 13:1178–1185

Fujita R, Hanauer A, Sirugo G, Heilig R, Mandel J-L (1990) Additional polymorphisms at marker loci *D9S5* and *D9S15* generate extended haplotypes in linkage disequilibrium with Friedreich ataxia. Proc Natl Acad Sci USA 87: 1796–1800

Fujiwara TM, Morgan K, Schwartz RH, Doherty RA, Miller SR, Klinger K, Staislovitis P, et al (1989) Genealogical analysis of cystic fibrosis families and chromosome 7q RFLP haplotypes in the Hutterite brethren. Am J Hum Genet 44:327–337

Germino GG, Weinstat-Saslow D, Himmelbauer H, Gillespie GAJ, Somlo S, Wirth B, Barton N, et al (1992) The gene for autosomal dominant polycystic kidney disease lies in a 750-kb CpG-rich region. Genomics 13:144–151

Harley HG, Brook JD, Floyd J, Rundle SA, Crow S, Walsh KV, Thibault M-C, et al (1991) Detection of linkage disequilibrium between the myotonic dystrophy locus and a new polymorphic DNA marker. Am J Hum Genet 49:68–75

Harris PC, Thomas S, Ratcliffe PJ, Breunig MH, Cotto E, Lopez-Larrea C (1991) Rapid genetic analysis of families with polycystic kidney disease 1 by means of a microsatellite marker. Lancet 338:1484–1487

Hill WG (1975) Linkage disequilibrium among multiple neutral alleles produced by mutation in finite population. Theoret Pop Biol 8:117–126

Hill WG, Robertson A (1968) Linkage disequilibrium in finite populations. Theor Appl Genet 38:226–231

Huntington's Disease Collaborative Research Group (1993) A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. Cell 72:971–983

Kerem B-S, Rommens JM, Buchanan JA, Markiewicz D, Cox TK, Chakravarti A, Buchwald M, et al (1989) Identification of the cystic fibrosis gene: genetic analysis. Science 245:1073–1080

McAlpine PJ, Shows TB, Boucheix C, Huebner M, Anderson WA (1991) The 1991 catalog of mapped genes and report of the nomenclature committee. Cytogenet Cell Genet 58:5–102

Navon R, Proia RL (1991) Tay-Sachs disease in Moroccan Jews: deletion of a phenylalanine in the α-subunit of β-hexosaminidase. Am J Hum Genet 48:412–419

Ozelius LJ, Kramer PL, de Leon D, Risch N, Bressman SB, Schuback DE, Brin MF, et al (1992) Strong allelic association between the torsion dystonia gene (*DYT1*) and loci on chromosome 9q34 in Ashkenazi Jews. Am J Hum Genet 50:619–628

Pandolfo M, Sirugo G, Antonelli A, Weitnauer L, Ferretti L, Leone M, Dones I, et al (1990) Friedreich ataxia in Italian families: genetic homogeneity and linkage disequilibrium with the marker loci D9S5 and D9S15. Am J Hum Genet 47:228–235

Pras E, Aksentijevich I, Gruberg L, Balow JE Jr, Prosen L, Dean M, Steinberg AD, et al (1992) Mapping of a gene causing familial Mediterranean fever to the short arm of chromosome 16. N Engl J Med 326:1509–1513

Reeders ST, Hildebrand CE, Sutherland GR (1991) Report of the committee on the genetic constitution of chromosome 16. Cytogenet Cell Genet 58:643–685

Roth C (ed) (1972) Encyclopedia Judaica. Kater, Jerusalem

Sankila E-M, de la Chapelle A, Karna J, Forsius H, Frants R, Ericksson A (1987) Choroideremia: close linkage to DXYS1 and DXYS12 demonstrated by segregation analysis and historical genealogical evidence. Clin Genet 31:315–322

Shohat M, Bu X, Shohat T, Fischel-Ghodsian N, Magal N, Nakamura Y, Schwabe AD, et al (1992) The gene for familial Mediterranean fever in both Armenians and non-Ashkenazi Jews is linked to the α-globin complex on 16p: evidence for locus homogeneity. Am J Hum Genet 51:1349–1354

Sirugo G, Keats B, Fujita R, Duclos F, Purohit K, Koenig M, Mandel JL (1992) Friedreich ataxia in Louisiana Acadians: demonstration of a founder effect by analysis of microsatellite-generated extended haplotypes. Am J Hum Genet 50:559–566

Sohar E, Gafni J, Pras M, Heller H (1967) Familial Mediterranean fever: a survey of 470 cases and review of the literature. Am J Med 43:227–253

Sohar E, Pras M, Heller J, Heller H (1961) Genetics of familial Mediterranean fever. Arch Int Med 107:529–538

Thompson AD, Shen Y, Holman K, Sutherland GR, Callen DF, Richards RI (1992) Isolation and characterization of $(AC)_n$ microsatellite genetic markers from human chromosome 16. Genomics 13:402–408

Thompson EA, Deeb S, Walker D, Motulsky AG (1988) The detection of linkage disequilibrium between closely linked markers: RFLPs at the AI-CIII apolipoprotein genes. Am J Hum Genet 42:113–124

Weir BS, Cockerham CC (1978) Testing hypotheses about linkage disequilibrium with multiple alleles. Genetics 88:633–642

Shen, Y., Holman, K., Doggett, N.A., Callen, D.F., Sutherland, G.R. and Richards, R.I. (1993). Five dinucleotide repeat polymorphisms on human chromosome 16q24.2-q24.3. *Human Molecular Genetics*, 2(9), 1504.

Shen, Y., Holman, K., Doggett, N.A., Callen, D.F., Sutherland, G.R. and Richards, R.I. (1993). Six dinucleotide repeat polymorphisms on human chromosome 16q12.1 – q24.1. *Human Molecular Genetics*, 2(9), 1505.

Shen, Y., Holman, K., Doggett, N.A., Callen, D.F., Sutherland, G.R. and Richards, R.I. (1993). Three dinucleotide repeat polymorphisms on human chromosome 16p 13.11–p13.3. *Human Molecular Genetics*, 2(9), 1506.

Shen, Y., Holman, K., Doggett, N.A., Callen, D.F., Sutherland, G.R. and Richards, R.I. (1993). Four dinucleotide repeat polymorphisms on human chromosome 16. *Human Molecular Genetics*, 2(10), 1745.

Shen, Y., Holman, K., Doggett, N.A., Callen, D.F., Sutherland, G.R. and Richards, R.I. (1994). Dinucleotide repeat polumorphisms at the D16S525, D16S359. D16S531 and D16S522 loci. *Human Molecular Genetics*, 3(1), 210.

Callen, D.F., Eyre, H., Lane, S. and Shen, Y. et al. (1993). High resolution mapping of interstitial long arm deletions of chromosome 16: relationship to phenotype. *Journal of Medical Genetics,* 30(10), 828-832.

*cpub2693/21 February, 1994*

# A PCR-BASED GENETIC LINKAGE MAP OF HUMAN CHROMOSOME 16

Shen, Y., Kozman, H.M., Thompson, A., Phillips, H.A., Holman, K., Nancarrow, J., Lane, S., Chen, L-Z., Apostolou, S., Doggett, N.*, Callen, D.F., Mulley, J.C., Sutherland, G.R. and Richards, R.I.**

Centre for Medical Genetics,
Department of Cytogenetics and Molecular Genetics,
Women's and Children's Hospital,
North Adelaide, SA 5006, Australia

* Los Alamos National Laboratory, Los Alamos, NM, USA

** To whom correspondance should be addressed :

Dr. R.I. Richards,
Centre for Medical Genetics,
Department of Cytogenetics and Molecular Genetics,
Women's and Children's Hospital,
North Adelaide, SA 5006
Australia
Telephone : 618-204 7333
Fax : 618-204 7342

Subject Category : Linkage and genetic maps.

Running title : Human chromosome 16 : PCR-based genetic map.

**ABSTRACT**

A high-resolution cytogenetic-based physical map and genetic linkage map of human chromosome 16 has been developed based on 79 PCR-typable genetic markers. Markers were previously characterized polymorphic $(AC)_n$ repeats. Initially these were isolated randomly from a cosmid library of a human-rodent somatic cell hybrid (CY18) with chromosome 16 as the only human chromosome. Gaps in the linkage map were targetted by the isolation of markers from cosmids which physically map to the deficient regions. In combination, these two approaches led to the characterisation of 47 highly informative genetic markers spread along chromosome 16, some of which are closely linked to disease loci. In addition, 22 markers (D16S401-423) previously genetically mapped by Weissenbach et al., (1992) were also physically mapped. 10 markers characterized by other laboratories were physically mapped and genotyped on the CEPH families. These 32 markers were incorporated into the PCR-based map. 72 markers have heterozygosity >0.50 and 51 of these markers >0.70. By multipoint linkage analysis of these markers in the CEPH reference family panel, a framework genetic map and a comprehensive genetic map were constructed. The length of the sex-averaged framework genetic map is 152.1 cM. The average distance and the median distance between markers on this map are 3.2 cM and 2.7 cM, respectively, and the largest gap is 15.9 cM. These maps were anchored to the high-resolution cytogenetic map (on average 1.5 Mb per interval) by a panel of 67 breakpoint hybrid cell lines which divide human chromosome 16 into 66 intervals. Together these integrated genetic and physical maps of human chromosome 16 provide the basis for the localisation, and ultimately the isolation of disease genes which map to this chromosome.

## INTRODUCTION

The molecular cartography of individual human chromosomes can be approached from a variety of genetic and physical perspectives. The sex-averaged genetic length of chromosome 16 has been estimated to be about 145 cM, which accounts for approximately 3% of the human genome or an estimated 98 megabases of DNA (Morton, 1991). The expected number of the genes on chromosome 16 is 1500 - 3000 (Mulley and Sutherland, 1993). To date 38 cloned genes and 15 cDNA sequences have been mapped on chromosome 16 (Whitmore et al., 1993) as well as several (as yet uncloned) disease genes (Mulley and Sutherland, 1993). Chromosome 16 contains four well characterized fragile sites *FRA16A* (16p13.11), *FRA16E* (16p12.1), *FRA16B* (16q21-q22.1) and *FRA16D* (16q23.1) (Sutherland, 1993).

One of the goals of the Human Genome Project is the construction of genetic maps of all human chromosomes with highly polymorphic markers spaced an average of 2 to 5 centimorgans apart. A number of genetic maps have been constructed for human chromosome 16 (Donis-Keller et al., 1987; Keith et al., 1990; Julier et al., 1990; NIH/CEPH Collaborative mapping group (1992); Kozman et al., 1993;). These maps were based mainly on RFLP markers which are less efficient for localization of disease genes and for refining linkage distance for positional cloning than highly polymorphic PCR based markers. Therefore, there was a need for more highly informative markers placed on high density cytogenetic-based physical and genetic linkage maps of chromosome 16.

Simple tandem repeats (STR) are frequently highly polymorphic, of wide distribution in the human genome, and can be easily analysed using PCR techniques (Weber et al., 1989; Litt et al., 1989). These characteristics make them ideal genetic markers for construction of a genetic linkage map. In this report we describe a framework genetic linkage map and a comprehensive genetic linkage map of chromosome 16 composed of 79 AC repeat markers.

## MATERIALS AND METHODS

### PCR analysis of AC repeat polymorphisms

The isolation and characterization of all AC repeat polymorphisms used in this study have been reported in brief previously (references given in Table 1). Conditions for polymerase chain reaction analysis of AC repeat copy number polymorphism were also as previously described (Shen et al., 1991).

### Panel of Chromosome 16 Somatic Cell Hybrids

The panel of human-mouse somatic cell hybrids includes previously reported hybrids (Callen et al., 1992, Whitmore et al., 1993) and an additional hybrid CY200. This hybrid was constructed from the fibroblast cell line 46,XX,t(11;16)(p11.2;p13.3). The hybrid retains the der(16)t(11;16) chromosome. Further evaluation of the hybrid CY18A, a hybrid which is a spontaneously arising derivative of CY18, established that two regions of chromosome 16 were present. These were a small region in q13, and a region in q24 containing the APRT gene. The hybrid panel containing 67 breakpoint cell lines was utilized in the physical mapping of STR markers on chromosome 16. The hybrid cell lines are depicted in Fig.1.

**Physical Mapping and Characterization of STR Markers**

The PCR amplifications for physical mapping were performed by using previously described methods (Richards et al., 1991b; Thompson et al., 1992). The markers isolated randomly from the chromosome 16 cosmid library were mapped to a specific breakpoint interval by virtue of the presence or absence of the PCR product using somatic cell hybrid DNAs as templates in the PCR. Allele frequencies and expected heterozygosities were determined from 50 - 130 unrelated Caucasians and/or CEPH parents.

**Genotype Data and Linkage Analysis**

Genotyping of a total of 81 markers was performed on the CEPH (Centre d'Etude du Polymorphisme Humain, Paris) reference panel as defined in Dausset et al. (1990).

55 markers were genotyped in our laboratory. Of these, 39 markers were genotyped in 40 CEPH families and 16 markers were genotyped in the 8 same CEPH families (102, 884, 1331, 1332, 1347, 1362, 1413, 1416) as used by Weissenbach et al. (1992). The genotypes of markers D16S401-423 were from version 6 of the CEPH database (contributed by Weissenbach et al., 1992).

Multipoint linkage analysis was carried out using a Sun SPARC station IPC. Map distance and marker order were determined by likelihood analysis using CRIMAP (Version 2.4) (Lander and Green, 1987). The details of multipoint linkage analysis were the same as described by Kozman et al. (1993).

**RESULTS**

**Isolation of $(AC)_n$ STR markers**

All of the AC repeat polymorphisms used in the construction of the PCR-based genetic linkage map have been previously described in brief (see

Table 1). The detailed statistics of the approaches used to obtain these markers are given here in order to indicate their utility.

Two approaches were used to isolate simple tandem repeats. Firstly a random isolation approach was exploited to generate STRs from the chromosome 16 cosmid library (Stallings et al., 1990). The cosmid library was constructed from the human mouse somatic cell hybrid CY18 (Callen, 1986), which contains chromosome 16 as the only human chromosome. The method used for random isolation was the same as previously described by Thompson et al. (1992).

110 $(AC)_n$ positive subclones in total (including those of Thompson et al., 1992) were sequenced. 10 clones were independently isolated twice from the chromosome 16 cosmid library. Of the 48 $(AC)_n$ repeats for which oligo primers were designed, 24 were characterized and physically mapped to the breakpoint intervals of chromosome 16, 14 were mouse $(AC)_n$ repeats due to the 7% mouse DNA contained in the chromosome 16 cosmid library (Stallings et al., 1990). 10 failed to give satisfactory results with either mouse or human genomic DNA. 57 clones were discarded for assorted reasons including 1) $(AC)_n$ repeats were too short ($n < 16$) or absent from available sequence; 2) $(AC)_n$ repeats were flanked by repetitive sequences; 3) clones were too close to the cloning site to design primers. 17 $(AC)_n$ repeat markers ascertained via this approach were used in the construction of the genetic linkage map and are given in Table 1.

After the $(AC)_n$ repeats isolated from this random approach were mapped to 16 breakpoint intervals of chromosome 16, a second approach was used to generate additional repeat markers with which to fill in the deficient intervals or to isolate more repeat markers in the regions of particular interest (fragile sites *FRA16A* and *FRA16B*, disease gene regions of PKD1, FMF and Batten disease).

86 cosmid clones were selected from the specific regions of chromosome 16 (Callen et al., 1992). 74 cosmid clones were $(AC)_n$ positive and subcloned into M13. 70 clones were sequenced. Oligo primers were designed for 35 $(AC)_n$ repeats. In two instances, identical $(AC)_n$ repeats were isolated from cosmids which were subsequently shown to overlap. The rest of the clones were discarded for the same sort of reasons as above described. 30 of these $(AC)_n$ repeat markers used for construction of the genetic linkage map are given in Table 1.

**Physical Mapping**

47 $(AC)_n$ repeat markers isolated in our laboratory were physically mapped using the hybrid panel. In addition, 22 markers (D16S401-423) from Weissenbach et al. (1992), 5 markers (D16S261, 265, 266, 397, 398) from Weber et al. (1990) and personal communication (1991), 2 markers (D16S310 and D16S313) from Hudson et al. (1992), D16S283 (SM7) from Harris et al. (1991), D16S285 from Konradi (1991) and SPN from Regaev et al. (1992) were also physically mapped using the hybrid panel. 79 STR markers mapped to 35 cytogenetic intervals (Fig.1.).

**Detection of Genotyping Inconsistencies**

The CHROMPIC option of CRI-MAP was used to detect double recombinants. When the double recombinations were detected, the original data were reexamined and then these DNA samples were regenotyped. False double recombination events were eliminated by this procedure. After regenotyping the original DNA samples (some samples were regenotyped twice or more), 13 double recombinants within a 15 cM region (9 of them within 10 cM) still remained. These double recombinants were included in the linkage analysis.

A total of 34,904 parent-offspring transfers of alleles were examined through the genotyping of 55 STR markers in the 40 CEPH reference families.

14 mutation events which did not show Mendelian inheritance were detected using DNAs from transformed lymphoblastoid cell lines. The mutation rate for these 55 chromosome 16 STR markers is $4.0 \times 10^{-4}$ per locus per gamete per generation.

Null alleles were found in 7 markers (Callen et al. 1993). These were recognised by the apparent noninheritance of a parental allele by a sib.

**Framework Genetic Map**

The framework map contains 48 markers (Figure 1), one of which is the VNTR marker (3'HVR, *D16S85*)(Kozman et al., 1993). *D16S85* and *D16S320* are two of the most informative markers and were used to initiate construction of the framework map. The 48 markers can be placed with likelihood odds of 1000:1 or greater in favour of one position over any other position on the map. The order of the framework loci established by linkage was compatible with their physical order independently determined on the cytogenetic map (Figure 1). The averaged distance and the median distance between markers in the sex-averaged map is 3.2 cM and 2.6 cM, respectively. The length of the sex-averaged map is 155.7 cM. The largest gap between loci is 16.1 cM.

**Comprehensive Genetic Map**

32 STR markers and one RFLP marker (*D16S44*)(Kozman et al., 1993), in addition to the 48 framework markers, were incorporated into the comprehensive map (Table 2). These 33 markers with an interval support of less than 3 (1000 : 1 odds) could not be uniquely ordered on the genetic map by linkage analysis (Figure 1). Many of these markers could be anchored, however, by physical mapping using the chromosome 16 hybrid panel. These 33 markers were placed within 16 different breakpoint intervals (Figure 1). The average interlocus distance in the comprehensive map is 1.9 cM. The largest gap between loci is 13.6 cM.

The distances between the loci in the sex-specific maps are shown in Table 2. The male map is 126.8 cM with an average interlocus distance of 1.6 cM, and the largest gap is 17.4 cM. The female map is 178.9 cM with an average interlocus distance of 2.2 cM, and the largest gap is 10.3 cM. The overall female to male ratio of recombination in the sex-specific maps is 1.4:1. On both distal regions of the chromosome 16 arms, the recombination rates on the male map is greater than the female map (2.9 fold between *D16S85* and *D16S406* on the distal short arm and 4.0 fold between *D16S449* and *D16S303* on the distal long arm). An excess of female over male recombination rates was seen in most genetic intervals throughout the remainder of this chromosome.

**DISCUSSION**

To isolate highly polymorphic markers for construction of a high-resolution genetic linkage map of human chromosome 16, random isolation was initially used. When it was found that the twenty four markers from random isolation were not uniformly distributed a targetted approach was adopted to isolate STR markers from specific regions of the chromosome otherwise deficient in them. In combination these two approaches led to the characterization of 47 highly informative markers. In addition 32 markers have been characterized from other laboratories. Of 79 STR markers, 72 markers have heterozygosity >0.50, and 51 have heterozygosity >0.70. Of 38 markers genotyped only in the 8 largest CEPH families, 16 (34% 16/47) were able to incorporated in the framework map and 22 (69% 22/32) were in the comprehensive map (Figure 1). The genotyping of markers in these 8 large families appears in some instances to be insufficient for the unique ordering of these markers by linkage analysis.

This PCR-based linkage map of human chromosome 16 was constructed from 79 STR markers, 1 VNTR marker, and 1 RFLP marker. These 2 non-STR markers

were chosen because they extend the map towards the telomeres. This map covers the entire length of chromosome 16, from *D16S85*, within 170-430 kb of the telomere on the short arm (Wilkie et al., 1991), to *D16S44*, within 230 kb of the telomere on the long arm (Kozman et al., 1993). The median distance and the average distance between markers on the framework map is 2.7 cM and 3.2 cM, respectively. In comparison with the averaged resolutions of other microsatellite linkage maps of 8 cM (chromosome 1, Engelstein et al., 1993), 15 cM (chromosome 4, Mills et al., 1992), 5.8cM (chromosome 11, Litt et al., 1993), 9.4 cM (chromosome 12, Dawson et al., 1993), 8.2 cM (chromosome 13, Petrukhin et al., 1993), 7.3 cM (chromosome 18, Straub et al., 1993), 8.5 cM (chromosome 20, Hazan et al., 1992), the resolution of this map is much higher.

The framework and comprehensive maps were anchored to the high-resolution cytogenetic map, which was divided into 66 breakpoint intervals (on average 1.5 Mb per interval) by a panel of 67 hybrids. The cytogenetic map is one of the most detailed map available for any of the autosomes (Callen et al., 1992). It is apparent that the combination of genetic linkage analysis and physical mapping can be extremely helpful in resolving locus order at the resolution of the comprehensive map. For example, the genetic location of marker *D16S452* is between *D16S287* and *D16S420* on the short arm, but the physical location is between hybrid breakpoints CY11 and CY180(P). Therefore the genetic location can be reduced to the interval between *D16S410* and *D16S420* (Figure 1). For eleven comprehensive map markers (*D16S94*, *D16S523*, *D16S405*, *D16S452*, *D16S383*, *D16S261*, *D16S411*, *D16S164*, *D16S347*, *D16S266* and *D16S363*), the distances of the location interval were narrowed by physical mapping (Figure 1).

The length of the centromeric region of chromosome 16 is 6-9 Mb. The genetic distance between the flanking markers (D16S300 and D16S285) is 1.0 cM. This demonstrates there is a possible suppression of recombination

over the centromere since the physical length is 6-9% of whole chromosome 16.

The sex-specific maps indicate the difference between the male and female recombination rates (Table 2). The female map is 1.4 times longer than the male map. Between the distal regions of chromosome arms, the greater recombination rate on the female map exhibits the same general trends as previous reports (Donis-Keller et al., 1987; Julier et al., 1990; Keith et al., 1990; Kozman et al., 1993). In contrast to the general phenomenon of a greater recombination rate in females, the recombination frequencies in two distal regions, *D16S406* to *D16S85* on the short arm and *D16S449* to *D16S303* on the long arm, shows greater recombination in males (Table 2). The genetic linkage maps of chromosome 16 published by Donis-Keller et al., 1987; Julier et al., 1990; Keith et al., 1990 and Kozman et al., 1993 also observed the same expansion of male genetic distance in telomeric regions.

The map presented here covers the entire chromosome 16 from *D16S85* to *D16S44*, and gives the genetic length of 126.8 cM in males, 178.9 cM in females, and an average of 151.1 cM in both sexes. By comparison, the genetic lengths in this report are in remarkable agreement with those published genetic and chiasma chromosome 16 maps (Table 3). The lengths of multipoint linkage maps are affected by the choice of mapping function and the frequency of errors (Bueton, 1991; Morton and Collins, 1990). Possible genotype errors detected as double recombinants using CHROMPIC were checked against the autoradiographs and regenotyping the original DNA samples. After this error checking, 13 double recombinants remained in the linkage data because they could not be excluded by regenotyping. The similar lengths of the genetic map reported here to those previously reported indicated that the length is apparently not affected by these double recombination events. 14 mutation events were detected as inconsistent genotypes as a result of genotyping the 55 STR markers, leading to a

mutation rate of $4.0 \times 10^{-4}$ per locus per gamete per generation. The mutation rates of simple tandem repeats estimated by Weber and Wong (1993) and Weissenbach et al. (1992) are $1.2 \times 10^{-3}$ and $1.0 \times 10^{-3}$, respectively. Thus the mutation rate reported here is 2.5-3 times less than those reported previously. If the 13 double recombinants were in fact new mutations, the mutation rate would be $7.7 \times 10^{-4}$ per locus per gamete per generation. Null alleles were found in 7 markers (Callen et al., 1993). In linkage analysis or in application to diagnosis, the presence of a segregating null allele will not corrupt the linkage data (Phillips et al., 1993) but could result in loss of information.

The distribution of simple tandem repeats is non-random on both the physical map and linkage maps (Figure 1) with a number of clusters and gaps. Nine breakpoint intervals on the physical map contain three or more STR markers. The average interlocus distance in seven clusters ranged from 0.10 to 0.88 cM. One interval on the short arm flanked by the CY165 and CY15 hybrid breakpoints contains 9 STR markers with average interlocus distance of 1.7 cM. Another interval on the distal long arm flanked by CY100, CY120 and CY18A(P) breakpoints contains 6 STR markers with average interlocus distance of 3.3 cM. The largest gap of 13.6 cM, between D16S405 and D16S414, is the only gap in excess of 10 cM. This gap is located at the interval flanked by CY19 and CY180(D) breakpoints on the physical map. This interval appears to be deficient in simple tandem repeats. Weissenbach et al. (1992) also observed non-uniform distribution of $(AC)_n$ repeats using random isolation from the whole human genome. There are several explanations of this non-random distribution (Weissenbach et al., 1992; Weissenbach, 1993) : 1) expansion of genetic distance of subtelomeric regions; 2) difficulties in PCR amplification of GC rich regions; 3) source of libraries for isolation of simple tandem repeats constructed using size fractionated fragments which were digested by specific restriction enzymes;

4) exclusion of less informative STRs and 5) primers could not be made for some STRs because of the limited flanking sequences. All of these could cause the observed clusters and gaps on the physical and genetic maps of chromosome 16.

These integrated genetic and physical maps of human chromosome 16 provide an efficient means for regional localization of genetic disorders located on chromosome 16, for detection of loss of heterozygosity in cancers and imprinting of chromosomes in inherited disorders, for evaluation of linkage disequilibrium and disease causing mutations, and for analysis of multifactorial diseases.

**ACKNOWLEDGMENTS**

This work was supported in part by the United States Department of Energy, contract number DEFG02-89ER 60863, the Australian National Health and Medical Research Council and by an International Research Scholars award from the Howard Hughes Medical Institute to Professor G. Sutherland. We thank Dr. J. Weber for provision of primer sequence data prior to its publication, Dr. K.A. Dyer for kindly providing the cell line used to construct the hybrid CY200. R.I.R. wishes to thank Shelley Richards for support and encouragement.

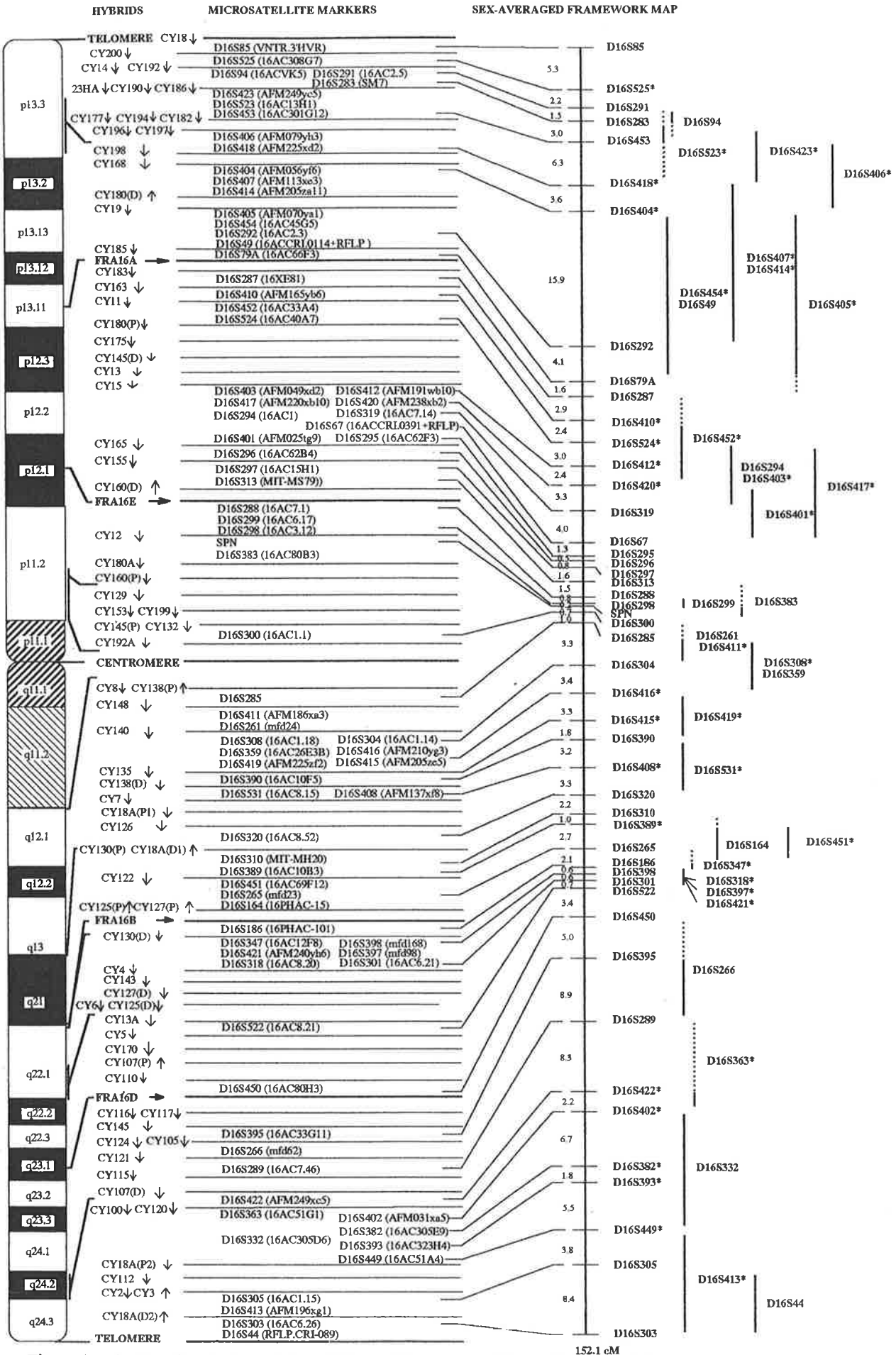



**Figure 1.** The Genetic and Physical Map of Human Chromosome 16

FIGURE CAPTION

Figure 1: The cytogenetic locations of the 81 markers and the sex-averaged framework map of human chromsome 16. The placements of the 33 comprehensive map markers are shown to the right (the dashed lines indicate the reduced interval distances determined by physical mapping using hybrid panel). The portion of chromosome 16 present in each human/mouse hybrid cell line is delineated by a horizontal line with an arrow indicating the direction of the retained portion of chromosome 16 from the breakpoint. Breakpoint of fragile sites are indicated by horizontal arrows. The 38 markers that were only genotyped on the 8 largest CEPH families are indicated by *.

**Table 2.** Distances (in cM) between the loci in the comprehensive map, including sex-specific and sex-average distances and likelihood support for order

| Locus Interval | Male | Female | Sex-average | Odds(1:) |
|---|---|---|---|---|
| D16S85 | 8.3 | 2.6 | 5.2 | $10^{11}$ |
| D16S525 | 4.0 | 0.0 | 2.0 | $10^{5}$ |
| D16S291 | 1.2 | 0.8 | 1.1 | $10^{6}$ |
| D16S283 | 0.0 | 1.0 | 0.4 | $10^{2}$ |
| D16S94 | 2.3 | 0.6 | 1.6 | $10^{3}$ |
| D16S523 | 2.6 | 1.3 | 1.8 | $10^{2}$ |
| D16S423 | 0.0 | 0.0 | 0.0 | 1.0 |
| D16S453 | 8.4 | 3.0 | 5.8 | $10^{15}$ |
| D16S406 | 0.0 | 0.0 | 0.0 | 1.0 |
| D16S418 | 2.7 | 5.6 | 4.1 | $10^{13}$ |
| D16S404 | 0.0 | 0.0 | 0.0 | 1.0 |
| D16S407 | 0.0 | 2.6 | 1.5 | $10^{3}$ |
| D16S414 | 17.4 | 10.3 | 13.6 | $10^{21}$ |
| D16S405 | 0.0 | 0.0 | 0.0 | 1.0 |
| D16S454 | 0.0 | 0.0 | 0.0 | 1.0 |
| D16S292 | 0.8 | 0.0 | 0.4 | 10 |
| D16S49 | 4.0 | 4.1 | 4.0 | $10^{15}$ |
| D16S79A | 2.1 | 0.8 | 1.6 | $10^{11}$ |
| D16S287 | 1.2 | 5.5 | 3.0 | $10^{10}$ |
| D16S410 | 0.5 | 2.8 | 2.0 | $10^{2}$ |
| D16S452 | 0.8 | 0.0 | 0.0 | 1.0 |
| D16S524 | 1.2 | 6.5 | 3.9 | $10^{7}$ |
| D16S412 | 0.0 | 3.2 | 1.3 | 2.0 |
| D16S403 | 0.0 | 0.0 | 0.0 | 1.0 |
| D16S417 | 0.0 | 0.0 | 0.0 | 1.0 |
| D16S294 | 0.0 | 1.1 | 0.6 | $10^{3}$ |
| D16S420 | 0.0 | 5.3 | 2.4 | $10^{9}$ |
| D16S401 | 1.5 | 0.0 | 1.2 | $10^{2}$ |
| D16S319 | 2.5 | 5.3 | 3.7 | $10^{28}$ |
| D16S67 | 0.0 | 2.1 | 1.3 | $10^{8}$ |
| D16S295 | 1.9 | 0.0 | 0.5 | $10^{6}$ |
| D16S296 | 0.0 | 1.4 | 0.7 | $10^{14}$ |

Table 2. Continued (1)

| Locus Interval | Male | Female | Sex-average | Odds(1:) |
|---|---|---|---|---|
| D17S297 | 0.0 | 3.0 | 1.6 | $10^{24}$ |
| 16S313 | 0.3 | 2.6 | 1.4 | $10^{23}$ |
| D16S288 | 0.3 | 1.5 | 0.9 | $10^{11}$ |
| D16S299 | 0.3 | 0.0 | 0.3 | 1.6 |
| D16S298 | 0.0 | 0.9 | 0.3 | $10^{4}$ |
| SPN | 0.1 | 0.0 | 0.0 | 1.0 |
| D16S383 | 0.2 | 0.8 | 0.7 | $10^{5}$ |
| D16S300 | 0.0 | 1.9 | 1.0 | $10^{8}$ |
| D16S285 | 0.2 | 0.0 | 0.0 | 1.0 |
| D16S411 | 0.3 | 0.8 | 0.5 | $10^{3}$ |
| D16S261 | 0.3 | 7.3 | 3.9 | $10^{17}$ |
| D16S308 | 0.2 | 0.0 | 0.0 | 1.0 |
| D16S304 | 0.7 | 0.0 | 0.4 | 1.9 |
| D16S359 | 0.0 | 5.5 | 2.7 | $10^{8}$ |
| D16S416 | 1.5 | 2.8 | 2.2 | $10^{7}$ |
| D16S415 | 0.0 | 0.0 | 0.0 | 1.0 |
| D16S419 | 0.0 | 7.9 | 3.5 | $10^{18}$ |
| D16S390 | 0.5 | 4.9 | 2.7 | $10^{10}$ |
| D16S408 | 0.3 | 0.0 | 0.0 | 1.0 |
| D16S531 | 0.4 | 7.2 | 3.7 | $10^{18}$ |
| D16S320 | 0.0 | 4.8 | 2.2 | $10^{17}$ |
| D16S310 | 0.0 | 1.9 | 1.0 | $10^{3}$ |
| D16S389 | 0.0 | 2.4 | 1.5 | $10^{4}$ |
| D16S451 | 2.4 | 0.3 | 1.3 | $10^{3}$ |
| D16S265 | 0.0 | 0.0 | 0.0 | 1.0 |
| D16S164 | 2.4 | 2.1 | 2.1 | $10^{14}$ |
| D16S186 | 0.0 | 0.0 | 0.0 | 1.0 |
| D16S347 | 0.0 | 1.0 | 0.6 | $10^{5}$ |
| D16S421 | 0.0 | 0.0 | 0.0 | 1.0 |
| D16S398 | 0.0 | 0.3 | 0.0 | 1.0 |
| D16S318 | 0.0 | 0.3 | 0.0 | 1.0 |
| D16S397 | 0.0 | 0.3 | 0.6 | 1.9 |
| D16S301 | 1.1 | 0.4 | 0.7 | $10^{5}$ |

**Table 2.** Continued (2)

| Locus Interval | Male | Female | Sex-average | Odds(1:) |
|---|---|---|---|---|
| D16S522 | 4.3 | 1.3 | 3.2 | $10^{15}$ |
| D16S450 | 5.1 | 6.9 | 5.3 | $10^{15}$ |
| D16S395 | 0.0 | 3.0 | 2.1 | $10^{2}$ |
| D16S266 | 3.5 | 10.6 | 7.1 | $10^{11}$ |
| D16S289 | 7.6 | 6.6 | 7.1 | $10^{19}$ |
| D16S422 | 0.0 | 0.0 | 0.0 | 1.0 |
| D16S363 | 1.7 | 2.2 | 1.9 | $10^{7}$ |
| D16S402 | 5.1 | 6.8 | 5.9 | $10^{18}$ |
| D16S392 | 0.0 | 4.2 | 2.0 | $10^{2}$ |
| D16S332 | 0.0 | 0.0 | 0.0 | 1.0 |
| D16S393 | 3.8 | 5.2 | 5.3 | $10^{16}$ |
| D16S449 | 9.8 | 1.1 | 4.5 | $10^{11}$ |
| D16S305 | 0.0 | 3.7 | 1.1 | $10^{2}$ |
| D16S413 | 10.8 | 0.4 | 6.3 | $10^{11}$ |
| D16S44 | 0.0 | 0.0 | 0.0 | 1.0 |
| D16S303 | | | | |

**Table 1.** The Simple Tandem Repeat Polymorphisms of Human Chromosome 16

| Locus | Clone name | Polymorphism Type | No. Alleles (Size Range) | Heter | Genotype### 133101 | Genotype### 133102 | Reference |
|---|---|---|---|---|---|---|---|
| D16S525 | 308G7** | (GT)10(GA)21(GT)21 | 17(143-175) | 0.91 | 159,161 | 149,165 | 37 |
| D16S291 | AC2.5* | (GT)25 | 9(154-170) | 0.79 | 162,164 | 156,156 | 41 |
| D16S283 | SM7 | (AC)19 | 11(81-107) | 0.65 | | | 9 |
| D16S94 | VK5B** | (GT)20 | 5(82-90) | 0.51 | 86,88 | 86,86 | 1 |
| D16S523 | 13H1** | (GT)19+(GT)3 | 5(77-87) | 0.68 | 77,77 | 77,85 | 36 |
| D16S423 | AFM249yc5# | (AC)17 | 10(121-139) | 0.73 | | | 45 |
| D16S453 | 301G12** | (GT)11+(GT)4 | 2(125-127) | 0.45 | 127,127 | 125,127 | 35 |
| D16S406 | AFM079yh3# | (AC)22+(AC)4 | 10( ND) | 0.81 | | | 45 |
| D16S418 | AFM225xd2# | (AC)24 | 10(166-188) | 0.82 | | | 45 |

* randomly isolated from chromosome 16 cosmid library.
** isolated from cosmid clones mapped in the specific regions.
*** provided by Dr. J.L.Weber (personal communication, 1991) were characterised, physically mapped and genotyped in our laboratory.
\# physically mapped in our laboratory.
## physically mapped and genotyped in our laboratory.
### genotype of individuals from CEPH reference pedigree #1331.

**Table 1.** Continued (1)

| Locus | Clone name | Polymorphism type | No. Alleles (Size Range) | Heter | Genotype### 133101 | Genotype### 133102 | Reference |
|---|---|---|---|---|---|---|---|
| D16S404 | AFM056yf6# | (CT)14+(AC)13 | 8(117-137) | 0.81 | | | 45 |
| D16S407 | AFM113xe3# | (AC)20 | 10(150-170 | 0.85 | | | 45 |
| D16S414 | AFM205za11# | (AC)19 | 6(152-161) | 0.61 | | | 45 |
| D16S405 | AFM070ya1# | (AC)25 | 8(116-150) | 0.76 | | | 45 |
| D16S454 | 45G5** | (GT)22 | 10(173-193) | 0.75 | 175,175 | 175,179 | 35 |
| D16S292 | AC2.3* | (GCT)7(GT)18 | 10(180-198) | 0.74 | 180,182 | 180,182 | 41 |
| D16S49 | CRI.0114** | (GT)6+(GT)3+(GT)14 | 7(108-132) | 0.68 | 110,130 | 108,130 | 35 |
| D16S79A | 66F3** | (GT)20 | 9(153-169) | 0.81 | 155,167 | 155,Null | 27 |
| D16S287 | 16EX81** | (GT)23 | 10(201-225) | 0.78 | 209, 217 | 213,217 | 25 |
| D16S410 | AFM165yb6# | (AC)13 | 4(134-140) | 0.55 | | | 45 |
| D16S452 | 33A4** | (GT)24 | 7(132-144) | 0.68 | 140,140 | 134,140 | 36 |
| D16S524 | 40A7** | (GT)19 | 11(143-169) | 0.76 | 163,169 | 143,155 | 36 |
| D16S412 | AFM191wb10# | (AC)4+(AC)21 | 7(101-125) | 0.75 | | | 45 |

Table 1. Continued (2)

| Locus | Clone name | Polymorphism type | No. Alleles (Size Range) | Heter | Genotype### 133101 | 133102 | Reference |
|---|---|---|---|---|---|---|---|
| D16S403 | AFM049xd2# | (AC)23 | 10(134-152) | 0.85 | | | 45 |
| D16S417 | AFM220xb10# | (AC)15+(AC)2 | 6(124-142) | 0.71 | | | 45 |
| D16S294 | AC1* | (GT)n | 2(138-140) | 0.49 | 136,136 | 136,136 | 41 |
| D16S420 | AFM238xb2# | (AC)14 | 8(179-201) | 0.81 | | | 45 |
| D16S401 | AFM025tg9# | (AC)19 | 8(ND) | 0.74 | | | 45 |
| D16S319 | AC7.14* | (GT)17+(T)8 | 6(145-159) | 0.52 | 149,157 | 149,149 | 32 |
| D16S67 | CRI.0391** | (GT)23 | 9(149-165) | 0.77 | 155, 161 | 153,163 | 4 |
| D16S295 | 62F3** | (TG)16 | 8(110-124) | 0.66 | 112,112 | 112,112 | 4 |
| D16S296 | 62B4** | (GT)16 | 7(151-163) | 0.75 | 155,159 | 155,161 | 4 |
| D16S297 | 15H1H** | (GT)7+(TG)15 | 7(232-244) | 0.73 | 232,242 | 236,242 | 4 |
| D16S313 | MIT-MS79## | (CA)20 | 10(139-157) | 0.57<br>0.77(CEPH) | 147,147 | 143,143 | 11 |
| D16S288 | AC7.1* | (GT)22 | 6(154-166) | 0.73 | 160,160 | 158,160 | 31 |

**Table 1.** Continued (3)

| Locus | Clone name | Polymorphism type | No. Alleles (Size Range) | Heter | Genotype### 133101 | 133102 | Reference |
|---|---|---|---|---|---|---|---|
| D16S299 | AC6.17* | (GT)19 | 8(126-140) | 0.72 | 132,138 | 136,136 | 41 |
| D16S298 | AC3.12* | (GT)8+(GT)24 | 11(172-192) | 0.79 | 182,190 | 180,182 | 41 |
| SPN## | | (TC)11(AC)13(GC)12 | 18(145-185) | 0.96 | | | 28 |
| D16S383 | 80B3** | (GT)15 | 5(148-156) | 0.45 | 148,152 | 152,152 | 36 |
| D16S300 | AC1.1* | (GA)12(GT)24(GA)5 | 7(163-175) | 0.61 | 165,167 | 167,167 | 41 |
| D16S285## | | (GT)n | 12(107-129) | 0.83 | | | 14 |
| D16S411 | AFM186xa3# | (AC)16 | 6(209-229) | 0.78 | | | 45 |
| D16S261 | MFD24## | (AC)7+(AC)14 | 6(88-100) | 0.66 | | | 43 |
| D16S308 | AC1.18* | (GT)17 | 10(135-165) | 0.77 | 155,159 | 157,157 | 41 |
| D16S304 | AC1.14* | (GT)20 | 12(132-154) | 0.60 | 140,152 | 140,140 | 41 |
| D16S359 | 26E3B** | (GT)12+(GC)4 | 5(88-104) | 0.42 | 92,96 | 92,92 | 37 |
| D16S416 | AFM210yg3# | (AC)3+(AC)14 | 4(217-223 | 0.42 | | | 45 |
| D16S415 | AFM205ze5# | (AC)22 | 8(208-234) | 0.72 | | | 45 |

Table 1. Continued (4)

| Locus | Clone name | Polymorphism type | No. Alleles (Size Range) | Heter | Genotype### 133101 | 133102 | Reference |
|---|---|---|---|---|---|---|---|
| D16S419 | AFM225zf2# | (AC)15(AG)12 | 9(146-164) | 0.75 | | | 45 |
| D16S390 | 10F5** | (GT)14+(GC)3+(GT)7 | 11(177-197) | 0.80 | 181,185 | 177,179 | 34 |
| D16S408 | AFM137xf8# | (AG)6+(AC)17 | 5(241-251) | 0.68 | | | 45 |
| D16S531 | 8.15* | (GT)31 | 17(111-149) | 0.86 | 125,149 | 119,123 | 37 |
| D16S320 | 8.52* | (GT)24 | 12(151-175) | 0.86 | 163,173 | 163,169 | 32 |
| D16S310 | MIT-MH20## | (ATAG)12 | 6(126-170) | 0.67 | 158,162 | 158,160 | 11 |
| D16S389 | 10B3** | (GT)16 | 11(88-114) | 0.77 | 98,110 | 98.100 | 34 |
| D16S451 | 69F12** | (GT)21 | 10(83-113) | 0.84 | 105,111 | 111,113 | 34 |
| D16S265 | MFD23## | (AC)20 | 12(89-117) | 0.75 | | | 43 |
| D16S164 | 16PHAC-15** | (GT)11+(GT)4(GA)9 | 5(169-179) | 0.38 | 169,169 | 169,169 | 26 |
| D16S186 | 16PHAC-101** | (GT)13 | 10(130-178) | 0.57 | 130,132 | 130,156 | 26 |
| D16S347 | 12F8** | (GT)18 | 9(186-204) | 0.76 | 186,190 | 186,190 | 34 |
| D16S421 | AFM240yh6# | (AC)18 | 5(206-216) | 0.56 | | | 45 |

**Table 1.** Continued (5)

| Locus | Clone name | Polymorphism type | No. Alleles (Size Range) | Heter | Genotype 133101 | Genotype 133102 | Reference |
|---|---|---|---|---|---|---|---|
| D16S398 | MFD168 | (GT)22 | 10(224-242) | 0.85 | 226,228 | 230,234 | *** |
| D16S318 | AC8.20* | (GT)22 | 6(134-152) | 0.54 | 136,144 | 144,146 | 32 |
| D16S397 | MFD98 | (GT)22 | 9(179-195) | 0.83 | 185,189 | 187,189 | *** |
| D16S301 | AC6.21* | (GT)19 | 6(142-152 | 0.64 | 146,146 | 146,146 | 41 |
| D16S522 | AC8.21** | (GT)21 | 8(103-119) | 0.69 | 105,109 | 105,113 | 37 |
| D16S450 | 80H3** | (GT)21 | 8(83-103) | 0.52 | 85,85 | 85,93 | 34 |
| D16S395 | 33G11** | (GT)16 | 9(98-124) | 0.69 | 98,98 | 98,114 | 34 |
| D16S266 | MFD62## | (AC)21 | 6(94-104) | 0.59 | | | 43 |
| D16S289 | AC7.46* | (GT)21 | 8(156-172) | 0.77 | 164,164 | 162,164 | 32 |
| D16S422 | AFM249xc5# | (AC)19+(AC)7 | 9(188-212) | 0.78 | | | 45 |
| D16S363 | 51G1** | (GT)20 | 9(242-258) | 0.78 | 246,256 | 248,254 | 33 |
| D16S402 | AFM031xa5# | (AC)23 | 12(161-187) | 0.86 | | | 45 |
| D16S382 | 305E9** | (GT)14+(TA)7+(T)10 | 9(113-151) | 0.78 | 119,121 | 115,119 | 33 |

Table 1. Continued (6)

| Locus | Clone name | Polymorphism type | No. Alleles (Size Range) | Heter | Genotype### 133101 | 133102 | Reference |
|---|---|---|---|---|---|---|---|
| D16S332 | 305D6** | (GT)13 | 3(122-128) | 0.53 | 122,122 | 122,126 | 33 |
| D16S393 | 323H4** | (GT)25 | 14(130-158) | 0.87 | 152,152 | 140,150 | 33 |
| D16S449 | 51A4** | (GT)21 | 9(183-201) | 0.85 | 193,197 | 195,195 | 33 |
| D16S305 | AC1.15* | (GT)27(AT)6+(GT)6 | 13(176-200) | 0.82 | 180,184 | 192,192 | 41 |
| D16S413 | AFM196xg1# | (A)15+(AC)24(AT)5 | 13(131-149) | 0.83 | | | 45 |
| D16S303 | AC6.26* | (GT)9+(GT)24+(GT)7 | 6101-115) | 0.43 | 101,103 | 109,113 | 41 |

**Table 3.** Estimates of Total Map Lengths for Human Chromosome 16

| Male (cM) | Female (cM) | Sex-Averaged (cM) | Reference |
|---|---|---|---|
| 164 | 237 | 195 | Donis-Keller et al. 1987 |
| 186 | 226 | | Julier et al. 1990 |
| 115 | 193 | 149 | Keith et al. 1990 |
| 120 | 193 | 157 | Morton et al. 1991 |
| 111 | 179 | 145 | Morton et al. 1991 (Chiasma data) |
| 133 | 197 | 162 | NIH/CEPH 1992 |
| | | 116* | Weissenbach et al. 1992 |
| 133 | 202 | 165 | Kozman et al. 1993 |
| 131 | 181 | 152 | This paper |

* This map did not extend the full length of human chromosome 16.

## References


1. Aksentijevich, I., Pras, E., Gruberg, L., Shen, Y., Holman, K., Helling, S., Prosin, L., Sutherland, G.R., Richards, R.I., Ramburg, M., Dean, M., Pras, M., Amos, C.I. and Kastner, D.L. (1993). Refined mapping of the gene causing familial mediterranean fever by linkage and homozygosity studies. *Am J Hum Genet* **53**: 451-461

2. Buetow, K.H., (1991). Influence of abberrant observations on high-resolution linkage analysis outcomes. *Am J Hum Genet* **49**: 985-994

3. Callen, D.F., Doggett, N.M., Stallings, R.L., Chen, L.Z., Whitmore, S.A., Land, S.A., Nanacarrow. J.K., Apostolou, S., Thompson, A.D., Lapsys, N.M., Eyre, H.J., Baker, E.G., Shen, Y., Holman, K., Phillips, H., Richards, R.I. and Sutherland, G.R. (1992). High-resolution Cytogenetic-based physical map of human chromosome 16. *Genomics* **13**: 1178-1185

4. Callen, D. F., Thompson, A. D., Shen, Y., Philips, H. A., Richards, R. I., Mulley, J. C. and Sutherland, G. R. (1993). Incidence and origin of "Null" alleles in the (AC)n microsatellite markers. *Am J. Hum Genet* **52:** 922-927

5. Dausset, J., Cann, H., Cohen, D., Lathrop, M., Lalouel, J-M. and White, R. (1990). Centre d'Etude du Polymorphisme Humain (CEPH): collaborative genetic mapping of the human genome. *Genomics* **6**: 575-577

6. Dawson, E., Shaikh, S,. Weber, J.L., Wang, Z., Weissenbach, J., Powell, J.F., and Gill, M. (1993). A continuous linkage map of 22 short tandem repeat polymorphisms on human chromosome 12. *Genomics* **17**: 245-248

7. Donis-Keller, H., Green, P., Helms, C., Cartinhour, S., Weiffenbach, B., Stephens, K., Keith, T.P., Bowden, D.W., Smith, D.R., Lander, E.S., Botstein, D., Akots, G. Rediker, K.S., Gravius, T., Brown, V.A., Rising, M.B., Parker, C., Powers, J.A., Watt, D.E., Kauffman, E.R., Bricker, A., Phipps, P., Muller-Kahle, H., Fulton, T.R., Ng, S., Schumm, J.W., Braman, J.C., Knowlton, R..G, Barker, D.F., Crooks, S.M., Lincoln, S.E., Daly, M.J. and Abrahamson, J. (1987). A Genetic linkage map of the human . *Cell* **51**:319-337

8. Engelstein, M., Hudson, T., Lane, J.M., Lee, M.K., Leverone, B., Landes, G.M., Peltonen, L., Weber, J.L. and Dracopoli, N.C. (1993). A PCR-based linkage map of human chromosome 1. *Genomics* **15**: 251-258

9. Harris, P.C., Thomas, S., Ratcliffe, P.J., Breuning, M.H., Coto, E. and Lopez -Larrea, C. (1991). Rapid genetic analysis of families with polycystic kidney disease 1 by means of a microsatellite marker. *Lancet* **338**: 1484-1487

10. Hazan, J., Dubay, C,. Pankowiak, M-P., Becuwe, N. and Weissenbach, J. (1992). A genetic linkage map of human chromosome 20 composed entirely of microsatellite markers. *Genomics* **12**: 183-189

11. Hudson, T.J., Engelstein, M., Lee, M.K., Ho, E.C., Rubenfield, M.J., Adams, C.P., Housman, D.E. and Dracopoli, N.C. (1992). Isolation and chromosomal assigment of 100 highly informative simple sequence repeat polymorphisms. *Genomics* **13**:

622-629

12. Julier, C., Nakamura, Y.A., Lathrop, M., O'Connell, P., Leppert, M., Mohandas, T., Lalouel, J-M. and White, R. (1990). A primary map of 24 loci on human chromosome 16. *Genomics* **6**: 419-427
13. Keith, T.P., Green, P., Reeders, S.T., Brown, V.A., Phipps, P., Bricker, A., Falls, K., Rediker, K.S., Powers, J.A., Hogan, C., Nelson, C., Knowlton, R. and Donis-Keller, H. (1990). Genetic linkage map of 46 DNA markers on human chromosome 16. *Proc Nalt Acad Sci USA* **87**: 5754-5758

14. Konradi, C., Ozelius, L., Yan, W., Gusella, J.F. and Breakefield, X.O. (1991). Dinucleotide repeat polymorphism (D16S285) on human chromosome 16. *Nucl Acid Res* **19**: 5449

15. Kozman, H.M., Philips, H.A., Callen, D.F., Sutherland, G.R. and Mulley, J.C. (1993). Integration of the cytogenetic and genetic linkage maps of human chromosome 16 using 50 physical entervals and 50 polymorphic loci. *Cytogenet and Cell Genet* **62**: 194-198

16. Lander, E.S., and Green, P. (1987). Construction of multilocus genetics linkage maps in human. *Proc Natl Acad Sci, USA* **84** 2363-2367

17. Litt, M., and Luty, J.A. (1989). A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *Am J Hum Genet* **44**: 397-401

18. Litt, M., Kramer, P., Hauge, X.Y., Weber, J.L., Wang, Z., Wilkie, P.J., Holt, M.S., Mishra, S. and Donis-Keller, H. (1993). A microsatellite-based index map of human chromosome 11. *Hum Mol Genet* **2**: 909-913

19. Mills, K.A., Buetow, K.H., Xu, Y., Weber, J.L., Altherr, M.R., Wasmuth, J.J. and Murray, J.C. (1992). Genetic and physical maps of human chromosome 4 based on dinucleotide repeats. *Genomics* **14**: 209-219

20. Morton, N.E. (1991). Parameters of the human genome. *Proc Natl Acad Sci USA* **88**: 7474-7476

21. Morton, N.E. and Collins, A. (1990). Standard maps of chromosome 10. *Ann Hum Genet* **54**: 235-251

22. Mulley, J.C. and Sutherland, G.R. (1993). Integrating maps of chromosome 16. *Current Opinion in Genetics and Development* **3**: 425-431

23. NIH/CEPH Collaborative Mapping Group (1992). A comprehensive genetic linkage map of the human genome. *Science* **258**: 67-162

24. Petrukhin, K.E., Speer, M.C., Cayanis, E., Bonaldo, M.F., Tantravahi, U., Soares, M.B., Fischer, S.G., Warburton, D., Gilliam, T.C. and Ott, J. (1993). A microsatellite genetic linkage map of human chromosome 13. *Genomics* **15**: 76-85

25. Phillips, H.A., Hyland, V.J., Holman, K., Callen, D.F., Richards, R.I. and Sutherland, G.R. (1991a) Dinucleotide repeat polymorphism at D16S287. *Nucl Acid Res* **19**: 6664

26. Phillips, H.A., Harris, P., Richards, R.I., Sutherland, G.R. and Mulley, J.C. (1991b). Dinucleotide repeat polymorphisms at the D16S164, D16S168, and D16S186 loci at 16q21-q22.1. *Nucl Acid Res* **19**: 6964

27. Phillips, H. A., Thompson, A. D., Kozman, H. M., Sutherland, G. R. and Mulley, J. C. (1993). A microsatellite marker whithin the duplicated D16S79 locus has a null allele: significance for linkage mapping. *Cytogenet Cell Genet* **63:** 131-132

28. Rogaev, E.I. and Keryanov, A.S. (1992). Unusual variability of the complex dinucleotide repeat block at the SPN locus. *Hum Mol Genet* **1**: 657

29. Richards, R.I., Holman, K., Lane, S., Sutherland, G.R. and Callen, D.F. (1991a). Human chromosome 16 physical map: mapping of somatic cell hybrids using multiplex PCR deletion analysis of sequence tagged sites. *Genomics* **10**: 1047-1052

30. Richards, R.I., Shen, Y., Holman, K., Kozman, H.M, Hyland, V.J., Mulley, J.C. and Sutherland, G.R. (1991b). Fragile X syndrome: diagnosis using highly polymorphic microsatellite markers. *Am J Hum Genet* **48**: 1051-1057

31. Shen, Y., Holman, K., Thompson, A., Kozman, H., Callen, D.F., Sutherland, G.R. and Richards, R. I. (1991). Dinucleotide repeat polymorphism at the D16S288 locus. *Nucl Acid Res* **19:** 5445

32. Shen, Y., Thompson, A., Holman, K., Callen, D.F., Sutherland, G. R. and Richards R. I. (1992). Four dinucleotide repeat polymorphisms on human chromosome 16 at D16S289, D16S318, D16S319 and D16S320. *Hum Mol Genet* **1:** 773

33. Shen, Y., Holman, K., Doggett, N.A., Callen, D.F., Sutherland, G.R. and Richards, R.I. (1993a). Five dinucleotide repeat polymorphisms on chromosome 16q24.2-q24.3. *Hum Mol Genet* **2**: 1504

34. Shen, Y., Holman, K., Doggett, N.A., Callen, D.F., Sutherland, G.R. and Richards, R.I. (1993b). Six dinucleotide repeat polymorphisms on chromosome 16q12.1-q24.1. *Hum Mol Genet* **2**: 1505

35. Shen, Y., Holman, K., Doggett, N.A., Callen, D.F., Sutherland, G.R. and Richards, R.I. (1993c). Three dinucleotide repeat polymorphisms on chromosome 16p13.11-p13.3. *Hum Mol Genet* **2**: 1506

36. Shen, Y., Holman, K., Doggett, N.A., Callen, D.F., Sutherland, G.R. and Richards, R.I. (1993d). Four dinucleotide repeat polymorphisms on chromosome 16. *Hum Mol Genet* **2**: 1745

37. Shen, Y., Holman, K., Doggett, N.A., Callen, D.F., Sutherland, G.R. and Richards, R.I. (1994). Dinucleotide repeat polymorphisms at the D16S525, D16S359, D16S531 and D16S522 loci. *Hum Mol Genet* in press

38. Stallings, R.L., Torney, D.C., Hildebrand, C.E., Longmire, J.L., Deaven, L.L., Jett, J.H., Doggett, N.A. and Moyzis, R.K. (1990). Physical mapping of human chromosomes by repetitive sequence finger printing. *Proc Natl Acad Sci USA* **87**: 6218-6222

39. Straub, R.E., Speer, M.C., Luo, Y., Rojas, K., Overhauser, J., Ott, J. and Gilliam, T.C. (1993). A microsatellite genetic linkage map of human chromosome 18. *Genomics*

**15**: 48-56

40. Sutherland, G.R. (1993). Human Fragile Sites. *In* "Genetic Maps. Locus maps of Complex Genomes. Book 5 : Human Maps" (S. J. O'Brien, Ed), Sixth Edition, pp.264-267, Cold Spring Harbor Laboratory Press, New York.

41. Thompson, A.D., Shen, Y., Holman, K., Sutherland, G.R., Callen, D.F. and Richards, R.I. (1992). Isolation and characterisation of (AC)n microsatellite genetic markers from human chromosome 16. *Genomics* **13**: 402-408

42. Weber, J.L. and May, P.E. (1989). Abundant calss of human DNA polymorphisms which can be typed using polymerase chain reaction. *Am J Hum Genet* **44**: 388-396

43. Weber, J.L., Kwitek, A.E. and May, P.E. (1990). Dinucleotide repeat polymorphisms at the D16S260, D16S261, D16S265, D16S266 and D16S267 loci. *Nucl Acid Res* **18**:4034

44. Weber,J.L. and Wong C. (1993) Mutation of human short tandem repeats. *Hum Mol Genet* **2**: 1123-1128

45. Weissenbach, J., Gyapay, G., Dib, C.S., Vignal, A., Morissette, J., Millasseau, P., Vayssiex, G. and Lathrop, M. (1992). A second-generation linkage map of the human genome. *Nature* **359**: 794-801

46. Weissenbach, J. (1993) Microsatellite polymorphisms and the genetic linkage map of the human genome. *Current Opinion in Genetics and Development* **3**: 414-417

47. Whitmore, S.A., Apostolou, S., Land, S., Nancarrow, J.K., Phillips, H.A., Richards, R.I., Sutherland, G.R. and Callen, D.F. (1993). Isolation and chracterisation of transcribed sequences from a chromosome 16 hn-cDNA library and the physical mapping of genes and transcribed sequence using a high resolution somatic cell panel of human chromosome 16. *Genomics* submitted

48. Wilkie, A.O.M., Higgs, D.R., Rack, K.A., Buckle, V.J., Spurr, N.K., Fischel-Ghodsian, N., Ceccherini, I., Brown, W.R.A .and Harris, P.C. (1991). Stable length polymorphism of up to 260kb at the tip of the short arm of human chromosome 16. *Cell* **64**: 595-606